人工智能很可能导致人类的永生或者灭绝，而这一切很可能在我们的有生之年发生。

THE REVOLUTION OF
ARTIFICIAL
INTELLIGENCE

Past Present and Future

人工智能革命

历史、当下与未来

王天一 | 著

北京时代华文书局

版权信息

书名：人工智能革命：历史、当下与未来

作者：王天一

出版社：北京时代华文书局

出版时间：2017年6月

ISBN:9787569915952

目录

前言

人工智能引发了对如何创建能够进行智能行为的计算机的广泛关注。多年来，缓慢却稳定进展逐渐使计算机在日常工作上变得更加“聪明”，研究界和行业的一系列突破近来更激发了对这一领域发展的势头和投资。

今天的人工智能仅限于狭隘的具体任务，并没有展现出如人类一般普适性的智慧。尽管如此，人工智能对世界的影响力依然不断增长。我们所看到的进步速度将对从医疗保健到形象和语音识别等领域产生不可估量的影响。在医疗保健方面，大量医疗计划将依靠人工智能来寻找医疗数据模式，帮助医生诊断疾病并提出治疗方案。在教育方面，人工智能有潜力帮助教师根据学生的需求定制教学。在交通领域，人工智能是自动驾驶技术的关键，这些无人操控的车辆与飞机可能会在未来几十年内改变全球物流系统。

人工智能带来的经济前景同样令人振奋：它无疑将重塑经济的面貌。根据埃森哲的研究报告估计，到2035年，人工智能可以使许多发达国家的年度经济增长率翻倍，并借此促进人与机器之间的新关系。该报告指出，业务中的人工智能将加强劳动者在推动业务增长方面的角色，进而提高劳动生产率。随着人工智能的成熟，它将潜在地成为近几十年来技术劳动生产率停滞和短缺的有力解决方案。

虽然许多人认为人工智能会取代人类，但人工智能令把人类的精力留在更杰出的工作上。创新的人工智能技术将使人们能够更有效地利用

自己的时间，做人类最好的工作——创造，想象和创新。即使在人工智能时代，成功和创造价值的关键因素仍然是采取“以人为本”的方针。

然而，像任何变革技术一样，人工智能带来了一些风险，并且从工作和经济到安全和监管问题的几个方面提出了复杂的政策挑战。人工智能系统也可以以惊人的方式表现，我们越来越依赖人工智能来提供决策和操作设备，这无疑增加了预测和控制复杂技术将如何行为的挑战。

无论如何，人工智能可以让人在全新的层面上工作，以推动增长和生产力。人工智能的核心任务不仅仅是消除重复的任务，而是把人放在中心，通过应用机器的能力来增加员工队伍，使人可以专注于更高价值的分析，决策和创新。

作者

第一章　用智慧再造智慧
——人类的终极梦想

世纪之交上映的三部曲电影《黑客帝国（Matrix）》，堪称人工智能题材中前无古人后无来者的里程碑式作品，其思想的深度与广度令后来者难以忘其项背。电影中，作为超级人工智能的Matrix将世界变成一个庞大的矩阵，支配这个矩阵运行的所有规律都在其掌握之下，人类反叛者尼奥（Neo）也不过是这个庞大棋局中的一个子。

设计师：你好，尼奥。

尼奥：你是谁？

设计师：我是设计师，我创造了Matrix.我一直在等你。我知道你有很多问题要问，虽然整个过程改变了你的意识，但你依然是不折不扣的人类。所以，我的一些回答你也许能明白，有些你也许不能明白。你的第一个问题也许是最关键的一个问题，同时你也许意识到或没有意识到它也是最无关紧要的问题。

尼奥：为什么我会在这里？

设计师：你的生命是Matrix固有程序中一个失衡因式的残留总和。你是一个偏差的偶然性，是尽管我竭尽全力，仍不能消除的影响数学精度和谐的一个偏差。尽管它不断地制造麻烦让我小心翼翼地处理它，但它并不是不可预测的，它仍然处于控制范围之内。它引导着你来到这里。

尼奥：你还没有回答我的问题。

设计师：很好。有意思，这要比其他的那些要快一点……Matrix比你想象的要老得多。我比较喜欢用一个完整偏差的出现到下一个完整偏差出现的方式来计算，这已经是第六个版本的Matrix.

尼奥：只可能有两种解释：没人告诉过我或是从来就没人知道。

设计师：正确。因为你无疑是在最简单化的因式里聚集并创造着偏差的系统化变动。

尼奥：选择。问题的关键是选择。

设计师：我设计的第一个Matrix非常完美，它简直就像是一件完美而卓越的艺术品。它的成功和失败都同样是史诗性的。它失败的必然性在我看来是每个人类固有的非完美性的结果。所以我根据你们人类的历史重新设计了Matrix，以便更准确地反映你们人类本性中多变的怪诞特质。可是我再次失败了。我终于了解到我得不到正确答案是因为它不需要太多的考虑，或是也许不需要考虑太多完美性的问题因素。答案最终被另一个指导性的程序偶然发现，这个程序原本是为了研究某些人类思维的。如果说我是Matrix之父，她无疑是Matrix之母。

尼奥：先知。

设计师：嗯。正如我所说的，她偶然发现了一个方法使得将近99.9%的试验体接受程序，给他们一个选择的机会，他们甚至只是仅仅意识到这个选择只是处于无意识的阶段。这个解决方案最初进行时，它无疑从基础上是有缺陷的，因而产生了相矛盾的系统偏差，如果不加以抑制就会威胁到系统本身。因此，那些拒绝程序的试验体，尽管只是少数，但如果不加以抑制，就会不断增加形成灾难的可能性。

今日世界，计算无处不在。

从结绳到算盘，从计算尺到集成芯片，计算的工具终于完成了从量

到质的飞跃。无所不能的计算如同《西游记》里的孙悟空，虽有万般变化，却仍然逃不脱逻辑规律这位如来佛祖的掌心。第一个发现这个秘密的人，便是英国数学家阿兰·图灵（Al-an Turing）。

图灵模型
——铺平理论道路

在1935年春天的剑桥大学国王学院，年仅23岁的图灵第一次接触到了德国数学家大卫·希尔伯特（David Hilbert）23个世纪问题中的第十问题：“能否通过机械化运算过程来判定整系数方程是否存在整数解？”

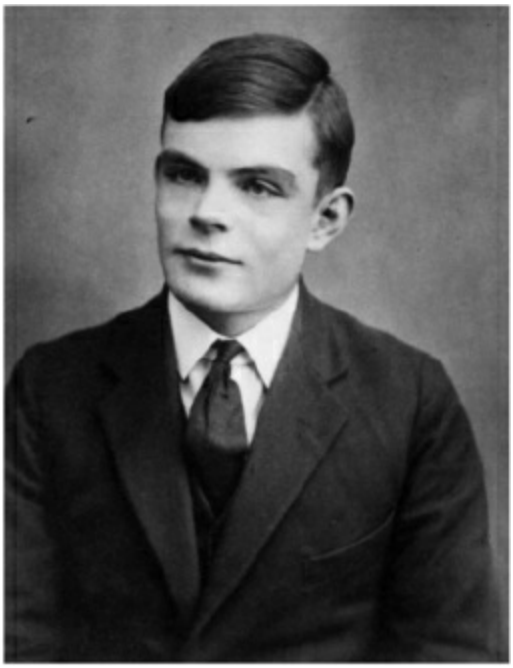

图1-1 阿兰·图灵

图灵清楚地意识到，解决这一问题的关键在于对“机械化运算”的严格定义。考究希尔伯特的原意，这个词大概意味着“依照一定的有限的步骤，无需计算者的灵感就能完成的计算”，这在没有电子计算机的当时已经称得上既富想象力又不失准确的定义。但图灵的想法更为单纯，机械计算就是一台机器可以完成的计算。用今天的术语来说，机械计算

的实质就是算法。

用机器计算的想法并不新鲜。17世纪，德国哲学家戈特弗里德·莱布尼兹（Gottfried Leibniz）就设想过用机械计算来代替哲学家的思考；两个世纪之后，计算机事业的先驱，英国工程师查尔斯·巴比奇（Charles Babbage）和他的红颜知己阿达·洛瓦莱斯（Ada Lovelace）就已经设计出了远远超越时代的“分析机”的原型。但图灵需要的机器跟各位先驱设想的机器稍有不同：它必须足够简单，可以用一目了然的逻辑公式描述它的行为；它又必须足够复杂，有潜力完成任何机械能完成的计算。这是一种能产生极端复杂行为的简单机器，这类机器在日后也被用他的名字命名为图灵机，以纪念这位伟大的先驱者。

1936年，图灵在伦敦权威的数学杂志上发表了划时代的重要论文《可计算数字及其在判断性问题中的应用》，首次提出了图灵机的概念。图灵机以天才的抽象性模拟了人脑的计算过程，将其还原为若干最基本的机械操作。对于人类而言，计算无非就是必备元素的集合：根据已有信息移动笔尖，在草稿纸上书写符号，指引书写的是一位数加法这些先验的规则，计算中涉及的进位操作则作为中间产物出现。在图灵机中，计算的每个必备元素都有其机械对应：笔被抽象为一个具有输入-输出功能的读写头，草稿纸被抽象为一条无限长度的纸带，先验的运算规则被抽象为读写头的内部状态转移表，一位数加法法则则被抽象为输入读写头的程序。

在运算过程中，图灵机的纸带被划分为小格，每格中只能有0和1两种符号，读写头则可以处于不同的状态中。在总共的有限个状态中，有一个特殊的“停机”状态。读写头一旦处于停机状态，就会停止运作；否则就会不停地运转下去。整台图灵机的核心在于读写头的状态转移表，它决定着读写头在读入来自纸带的一格信息后，其内部状态如何变化并将什么信息输出到纸带的格子上。

图灵机作为理论模型可谓“麻雀虽小五脏俱全”，它所能完成的任务

绝不像它看起来那么简单。只要有足够长的纸带和足够好的耐心，今天的计算机能做的计算，一台精心设计的图灵机同样能够完成。足够长的纸带可以模拟出足够大的寄存器、内存和硬盘；而在中央处理器的电路中，虽然所有可能的状态极多，但其数目终究是有限的，也就超不出图灵机的功能范畴。只不过这台图灵机的状态转移表将会有着超乎寻常的大小，以及通常超乎寻常的复杂程度，每次“读写内存”时，读写头都需要花长得令人咋舌的时间在纸带上来回奔波。

图灵机的出现本来是用于解决纯数学中的基础理论问题，可它却带来了意想不到的巨大收获：从理论上证明了研制通用数字计算机的可行性。图灵机的实际意义在于定义了数字计算机的计算能力：数字计算机能识别的语言属于递归可枚举的集合，它能计算的问题是部分递归函数的整数函数。

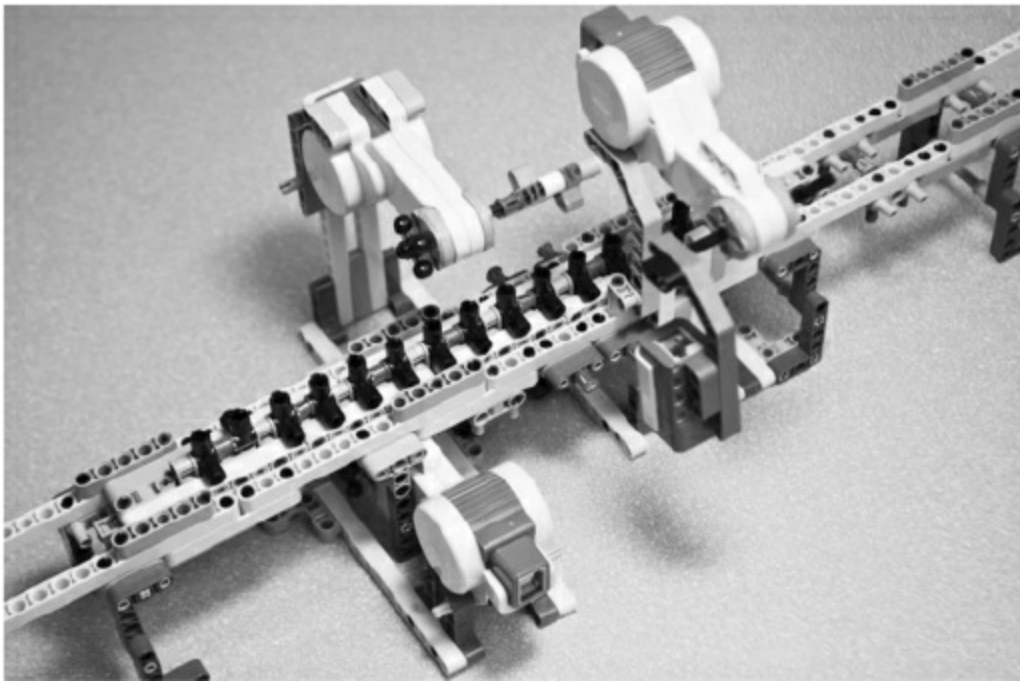

图1-2 用乐高积木搭建的图灵机

图灵机模型的成功丰富了图灵的想象力，他开始思考图灵机运算能力的极限。如果让图灵机拥有更多的纸带和对应的读写头，而纸带上也

不再限定两种符号，而是三种四种甚至更多种符号，图灵机就可以更快地实现预定任务。但从本质上来说，“升级”后的图灵机能完成的任务，原来的图灵机也能完成，差别只是出现在所需的时间资源上。换言之，这种“升级”并没有给可计算性带来任何质变，无论升级与否，能计算的问题仍然能计算，不能计算的问题也依然不能。显然，制约数字计算机的瓶颈并不在于性能指标。而在运行中遵循的逻辑规律。

自1940年起，图灵开始认真地思考机器是否能够具备类人的智能，而科学家敏锐的直觉使他马上意识到这个问题的关键其实并不在于如何打造强大的机器，而在于我们人类如何看待智能，即依据什么标准评价一台机器是否具备智能。于是在1950年，图灵发表了论文《计算机器与智能》，首次提出了对人工智能的评价准则，即大名鼎鼎的“图灵测试”。图灵测试是在测试者与被测试者（一个人和一台机器）隔开的情况下，由测试者通过一些装置向被测试者随意提问。如果经过5分钟的交流后，如果有超过30%的测试者不能区分出哪个是人、哪个是机器的回答，那么这台机器就通过了测试，并被认为具有人类水准的智能。

图1-3 电影《模仿游戏》海报

本质上说，图灵测试从行为主义的角度对智能进行了重新定义，它将智能等同于符号运算的智能表现，而忽略了实现这种符号智能表现的机器内涵。它将智能限定为对人类行为的模仿能力，而判断力、创造性等人类思想独有的特质则必然无法被纳入图灵测试的范畴。但无论图灵测试存在怎样的缺陷，它都是一项伟大的尝试。自此，人工智能具备了必要的理论基础，开始踏上科学舞台，并以其独特的魅力倾倒众生，带给人类关于自身、宇宙和未来的无尽思考。

当然，相较计算机专业领域的成就，图灵更加广为人知的事迹是他在第二次世界大战中为盟军胜利做出的卓越贡献。在德军凭借其密码机恩尼格玛（Enigma）实现了军事情报的保密传送，进而在欧洲战场占据先机的情形下，图灵毅然应征勤王，正式到“政府编码与密码学院”服役。在剑桥的布雷契莱庄园，图灵领导着由200余位数学家组成的智力大军，成功破解了恩尼格玛，使英军在战场上迅速扭转局势，决定了第二次世界大战的最终走向。虽是一介书生，却堪敌百万雄兵。改编这段经历的电影《模仿游戏》于2015年上映，在大银幕上再现了图灵的传奇人生。

然而，图灵对人类的贡献不仅在于破译了德军的密码，更在于破译了人类思维的一些重要秘密。

冯诺伊曼结构
——踏平技术坎坷

图灵奠定了计算机的理论基础，美国科学家约翰·冯诺依曼（Johnvon Neumann）则将图灵的理论物化成为实际的物理实体，成为了计算机硬件体系结构的奠基者。自第一台冯诺依曼计算机诞生以来，七十余年的时间悄然流逝，计算机的技术与性能在这期间都发生了翻天覆地的变化，不变的却是作为主流体系架构的冯诺依曼结构。

1936年，当意气风发的图灵来到美国的普林斯顿大学攻读数学博士学位时，在位于同一个城市的普林斯顿高等研究院，同是不世出的天才的冯诺依曼当时正在该研究院主持数学研究。他对图灵的才华赞叹不已，极力邀请图灵毕业后做他的研究助手，只可惜图灵心系剑桥，一心要回到母校任教，这不禁令冯诺依曼颇为惋惜。可令人惋惜的远不止此，如果当年两位科学奇才能够强强联手、通力合作，必将给数学和计算机科学等学科带来革命性的变革，只可惜这种美好的景象只能存在于想象之中。

冯诺伊曼于1903年出生于匈牙利布达佩斯的一个犹太人家庭，在量子力学、现代计算机、纯数学与应用数学、核武器和生化武器等诸多领域内都有杰出建树，是20世纪难得一见的百科全书式学术奇才。但在生涯早期，冯诺伊曼的运气实在欠佳。1931年，当冯诺依曼即将在希尔伯特世纪问题中第二问题上获得突破时，却突然得知奥地利逻辑学家库尔特·哥德尔（Kurt Godel）已经先他一步发表了哥德尔定理，着实令他郁闷不已。一气之下，冯诺伊曼转行研究量子力学。可就在他的量子力学研究即将结出硕果之际，另外一位天才物理学家英国人保罗·狄拉克

（Paul Dirac）又一次抢了他的风头，出版了奠基性的巨著《量子力学原理》，这比冯诺依曼的《量子力学的数学基础》整整早了两年。

图1-4 约翰·冯诺伊曼

接连受到两次打击之后，冯诺依曼开始把部分注意力从基础数学转向了工程应用领域，出众的才华终于结成硕果。在美国，冯诺依曼积极参与美国军方武器研制的过程。自此，他把自己巨大热情和天赋投入到计算机研制和运用的事业中。

经过长期的构思与讨论，冯诺依曼在火车上完成了离散变量自动电子计算机的设计。1945年6月30日，《关于离散变量自动电子计算机的草案》经油印复印，由莫尔学院限量发行，“冯诺依曼体系结构”就此诞生。诞生在列车上的卓越思想为电子计算机的逻辑结构设计奠定了基础，已成为计算机设计的基本原则。

冯诺依曼体系结构采用二进制代替十进制，因而完成了计算机从模拟到数字的转化。在硬件上，冯诺伊曼体系结构包括五大部分：运算

器、控制器、存储器、输入设备和输出设备，建立在硬件基础上的则是“存储程序”原理——使用同一个存储器，经由同一个总线传输，程序和数据统一存储同时在程序控制下自动工作。特别要指出的是，它的程序指令存储器和数据存储器是合并在一起的，程序指令存储地址和数据存储地址指向同一个存储器的不同物理位置。这是因为程序指令和数据都是用二进制码表示，且程序指令和被操作数据的地址又密切相关。

冯诺伊曼体系结构给计算机的性能带来了革命性突破：最早的计算机器仅仅内置固定用途的专用程序，因而只能实现特定的功能。如果想要改变此机器的程序，就必须更改线路、调整结构甚至重新设计机器。由于当年的计算机器并非今日可编程的计算机，彼时所谓的“重写程序”很可能指的是纸笔设计程序步骤，接下来制订工程细节，再施工将机器的电路配线或结构改变。冯诺伊曼体系结构的“存储程序”理念则将计算机的专用性拓展为通用性。借由创造一组指令集结构，并将所谓的运算转化成一串程序指令的执行细节，将指令转化为一种特别形态的静态资料。一台储存程序型电脑可轻易改变其程序，并在程序的控制下改变其运算内容。同时，“存储程序”理念也允许程序执行时自我修改程序的运算内容。事实上，这一理念的设计动机之一就是让程序自行增加内容或改变程序指令的内存位置，因为早期的设计都要使用者手动修改。但随着索引暂存器与间接位置存取变成硬件结构的必备机制后，本功能就不如以往重要了。而程序自我修改这项特色也被现代程序设计所扬弃，因为它会造成理解与除错的难度，且现代中央处理器的管线与快取机制会让降低这种功能的效率。

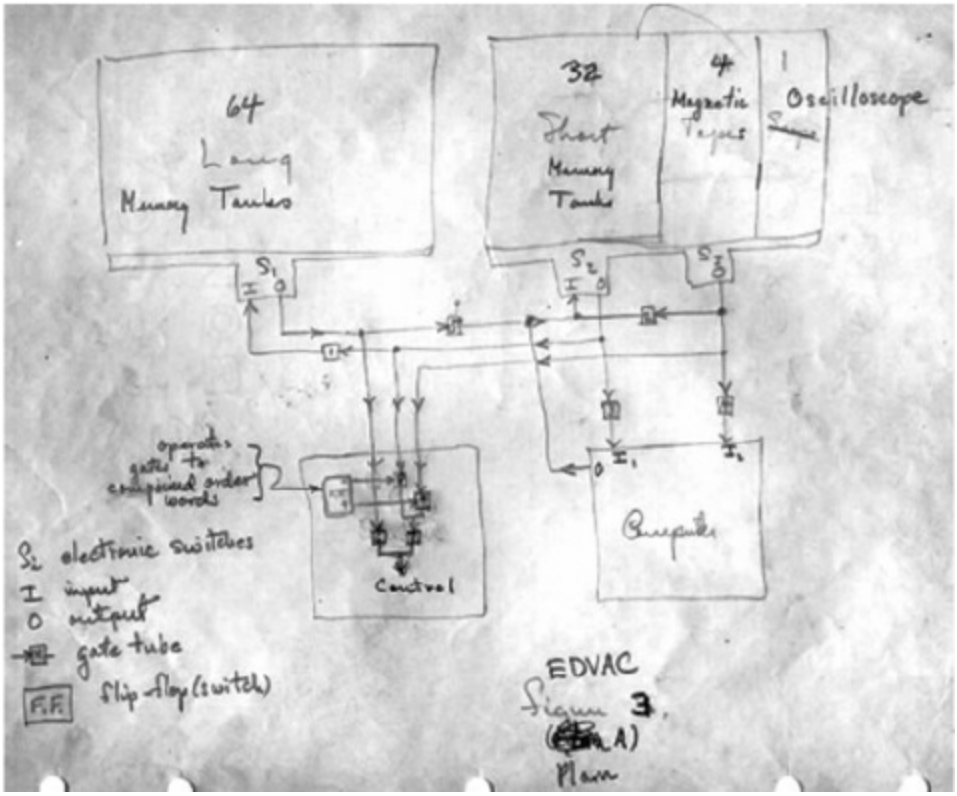

图1-5 冯诺伊曼设计的离散变量自动电子计算机草图

虽然踏平了计算机物理实现的技术坎坷，但冯诺依曼对建造计算机并没有太大的兴趣，他和图灵一样，更感兴趣的是计算机能够做什么。自然而然，冯诺依曼注意到机器的自动复制问题，并对生物世界的复制进行了同样深入的思考和比较。自动计算机能制造出和本身复杂度相当或者更高的后代吗？冯诺依曼解释说：“低级的‘复杂性’可能是退化性的，即每一个可以制造其他自动机的自动机只能产生较不复杂的自动机。然而，当复杂性超过某一个特定水平之后，如果对合成现象进行适当安排，就会发生爆炸性的变化。换句话说，自动机的合成可以通过这样一种方式进行：每个自动机有可能产生比自身更为复杂、更具潜力的自动机。”

可能连冯诺伊曼自己都没有意识到，他提出了一个重要的科学问题：能否创造出与人类智能具有相当水平的机器智能？但可惜的是，他的生命已经逐渐走向了尽头。1957年，53岁的冯诺伊曼因骨癌去世，而

骨癌的病因很可能源自大名鼎鼎的曼哈顿计划的核辐射。这位美国核武器的奠基人，临终时刻却只能在军方代表的监视下度过。而人类智能与机器智能的关系这个未竟的话题，也只有留待后来者去探索了。

最后，让我们用冯诺伊曼的一件轶事作为结束：

在一次晚会上，女主人勇敢地向冯诺伊曼提出一道数学题：

“相距一英里的两列火车在同一轨道上以每小时30英里的速度相向而行，这时栖在一列火车前面的一只苍蝇以每小时60英里的速度朝着另一列火车飞去，当它飞到另一列火车时又迅速地飞回来。它一直这样飞来飞去，直到两列火车不可避免地发生碰撞。请问这只苍蝇共飞了多少英里？”

大多数人，尤其是懂一点数学的人，都会先计算出苍蝇每一来回飞行的路程，再把这些结果累加起来。这一方法虽然直观，但涉及到无穷级数求和的问题，因而费时费力。更聪明的算法是首先计算出两列火车要经过多长时间才能碰撞，再用这个时间乘以苍蝇的飞行速度，清楚又简洁。

女主人话音刚落，冯诺依曼就脱口而出：“一英里。”

“天啊！你这么快就算出来了，”女主人惊呼，“大多数数学家都没能看出这里面的技巧，而是用无穷级数去计算，就不知道要算多久。”

“什么技巧？我也是用无穷级数算的。”冯诺依曼诧异地问。

达特茅斯之野望
——人工智能横空出世

当计算机的理论基础与工程技术全部成熟之后，人工智能的出现就可谓万事俱备，只欠东风。这一瓜熟蒂落的时刻的到来出现在1956年8月。彼时在美国达特茅斯学院（Dartmouth Col-lege）召开的学术会议，正可谓群贤毕至，少长咸集，汇聚了一大批未来学界的风云人物——包括1969年图灵奖获得者马文·闵斯基（Marvin Minsky）、1971年图灵奖获得者约翰·麦卡锡（John McCarthy）、1975年图灵奖获得者艾伦·纽埃尔（Allen Newell）、1975年图灵奖和1978年诺贝尔经济学奖获得者赫伯特·西蒙（Herbert Simon）在内的诸多科学家。在宁静的汉诺斯小镇，这些计算机大咖们正讨论着一个不食人间烟火的主题：用机器来模仿人类学习以及其他方面的智能。

这次会议最重要的成果就是确定了会议所讨论的研究内容的名称——由麦卡锡提出的人工智能（Artificial Intelligence），1956年也就成为了人工智能元年。正是在达特茅斯会议之后，人工智能进入了其发展的第一次黄金时期。

人工智能大展身手的第一个学科是纯数学学科，而最早取得突破的领域是使用计算机程序代替人类进行自动推理来证明数学定理。20年后分享图灵奖的纽埃尔和西蒙在达特茅斯会议上展示了人类历史上首个人工智能程序“逻辑理论家（Logic Theo-rist）"，它不仅证明出《数学原理》——阿尔弗雷德·怀特海（Alfred Whitehead）和贝特兰·罗素（Bertrand Russell）的三卷本数理逻辑巨著——中前52个定理中的38个，还给出了一些比罗素本人的证明更加简洁的解法，这甚至让罗素本

人兴奋不已。

机器定理证明的前进一发而不可收。1958年，美籍华人王浩在IBM704计算机上证明了《数学原理》中有关命题演算部分的全部220条定理，同年IBM公司还研制出了平面几何的定理证明程序。1959年，纽埃尔和西蒙又开发出一种不依赖于具体领域的通用问题求解器（General Problem Solver）。1961年，约翰·霍普金斯大学的美国学者詹姆斯·斯拉格（James Slagle）发表了一个符号积分程序（Symobolic Automatic Integrator），它能完成初等微积分中的不定积分式的计算。1963年，“逻辑理论家”也进化到能够证明《数学原理》的全部前52条定理。

“逻辑理论家”的出现在人工智能的历史上具有里程碑式的意义。这不仅因为它是第一个人工智能程序，更因为它颠覆了人们对计算机的印象：自1946年首台计算机诞生以来，计算机用来解决的都是诸如导弹弹道计算、核反应模拟这类具体数值的计算问题，抽象的、符号化的数学证明一直以来被认为超出了计算机的能力范围，“逻辑理论家”的出现显然颠覆了这固有的印象。其发明者西蒙曾评论道：“我们发明了具备抽象思考能力的程序……解释了合成的物质如何能够拥有人类的心智。”

遗憾的是，在关于机器与心智的判断上，西蒙和其他年少轻狂的人工智能科学家们都过于乐观了。1965年，人工智能在机器定理证明领域遭遇了滑铁卢，计算机推了数十万步也无法证明两个连续函数之和仍是连续函数。最糟糕的事情则发生在机器翻译领域，自然语言的理解与处理确实是人工智能中的硬骨头，但计算机在自然语言理解与翻译过程中表现之差也的确超乎了研究者的想象，一个最典型的例子就是下面这个著名的英语句子：

The spirit is willing but the flesh is weak.（心有余而力不足。）

可当计算机把这句话翻译成俄语再翻译回英语时，得到的结果可谓风马牛不相及：

The wine is good but the meet is spoiled.（酒是好的，肉变质了。）

骨感的现实打碎了丰满的理想，不仅让已被过誉的人工智能走下神坛，也给研究者们的头浇了一盆冷水。痛定思痛，他们开始思考如何突破这一瓶颈，美国计算机科学家爱德华·费根鲍姆（Edward Feigenbaum）正是人工智能新路的开拓者。受哲学家弗朗西斯·培根（Francis Bacon）“知识就是力量”著名论断的指引，费根鲍姆将视线从抽象的通用证明方法转移到具体的专业知识上，强调人工智能必须在知识的指导下实现，这催生了人工智能新领域——专家系统（Expert System）的诞生。

所谓的专家系统实质上是利用计算机基于已有的知识进行自动推理，从而从领域专家的角度解决实际问题。第一个实用的专家系统Dendral于1968年诞生，它可以根据质谱仪的数据推知物质的分子结构。在这个系统的影响下，各式各样的专家系统很快陆续涌现，形成了软件产业一个全新的分支：知识产业。1977年，在第五届国际人工智能大会上，费根鲍姆用术语“知识工程”为这个全新的领域命名。

可惜好景不长，在专家系统或知识工程获得大量的实践经验之后，其弊端也开始逐渐显现：它们的运作需要从外界获得大量知识的输入，而这样的输入工作是极其费时费力的，这就是知识获取的瓶颈。这个全新的棘手问题虽然没有催生新的“费根鲍姆”，却给人工智能这个学科带来了革命性的改变：它逐渐分化成了几大不同的学派，沿着不同的路径继续发展。

你方唱罢我登场
——三大流派竞风流

尽管传统的人工智能研究者也在奋力挣扎，但是他们不得不承认，如果采用完全不同的世界观，即让知识通过自下而上的方式涌现，而不是让专家们自上而下地设计出来，那么机器学习的问题其实可以得到很好地解决。这就好比我们教育小孩子，传统人工智能好像填鸭式教学，而新的方法则是启发式教学：让孩子自己来学。

事实上，在人工智能学术界，很早就有人提出过自下而上的涌现智能的方案，只不过它们从来没有引起大家的注意。一批人认为可以通过模拟大脑的结构（神经网络）来实现，而另一批人则认为可以从那些简单生物体与环境互动的模式中寻找答案。他们分别被称为连接学派和行为学派。与此相对，传统的人工智能则被统称为符号学派。自20世纪80年代到90年代的十年间，这三大学派形成了三足鼎立的局面。

符号主义学派的代表人物是达特茅斯会议的与会者之一麦卡锡，而他对人工智能的理解也代表了符号主义学派的见解：

“（人工智能）是关于如何制造智能机器，特别是智能的计算机程序的科学和工程。它与使用机器来理解人类智能密切相关，但人工智能的研究并不需要局限于生物学上可观察到的那些方法。”

图1-6 约翰·麦卡锡

符号主义学派认为人工智能源于数理逻辑，而数理逻辑才是智能行为的描述方式，用于机器定理证明的逻辑演绎系统事实上也继承了图灵测试的衣钵。该学派认为人类认知和思维的基本单元是符号，而认知过程就是对符号的逻辑运算，这样一来，人类抽象的逻辑思维就可以通过计算机中逻辑门的运算来模拟出来，进而实现机械化的人类认知，也就是人工智能。值得注意的是，麦卡锡着重强调人工智能的智能并不体现在真实的具体行为，而是体现在思维方式上，换言之，人类智能本身就能够被看成一类特殊的软件，至于运行它的硬件到底是碳基（人脑）还是硅基（计算机），反而没有那么重要了。

发明“逻辑理论家”的纽埃尔和西蒙则把麦卡锡的观点进一步推演为“物理符号系统假说”。该假说认为，任何能够将某些物理模式或符号转化成其他模式或符号的系统都有可能产生智能的行为，符号主义学派之名也由此而来。这种物理符号可以是人脑神经网络上的电脉冲信号，当然也可以是通过各种逻辑门产生的高低电平。在“物理符号系统假说”的支持下，符号学派把焦点集中在人类智能的高级行为，如推理、规划、知识表示等方面。这些工作曾在某些特定领域取得了空前的成

功。

图1-7 沃森在《危险游戏》节目中

1958年西蒙就曾预言，计算机会在10年内成为国际象棋世界冠军。这一天虽然在30年后才姗姗来迟，却也验证了西蒙的论断。2011年，由IBM公司制造的另一台超级计算机又创造了历史：在美国的电视节目《危险游戏（Jeopardy）》中，超级计算机沃森（Watson）通过处理自然语言线索，在涉及各个领域的知识问答上战胜了人类选手。沃森的胜利是人工智能界的一个标志性事件，它说明计算机不仅能在初始条件确定的棋盘博弈中获胜，在不存在初始条件与边界条件的开放世界中的表现同样不逊于人类，至少是在某些特定条件下。

一言以蔽之，人机大战是符号主义学派人工智能的标志性应用，但这样的“战争”对IBM公司市值的意义远大于对人工智能发展的意义。经过短暂的辉煌之后，符号主义学派也逐渐走向式微。

连接主义学派并不认为人工智能源于数理逻辑，也不认为智能的关键在于思维方式。这一学派把智能建立在神经生理学和认知科学的基础上，强调智能活动是由大量简单的单元通过复杂的相互连接后并行运行的结果。众所周知，人类的智慧主要来源于大脑的活动，而大脑则是由

一万亿个神经元细胞通过错综复杂的通路相互连接形成的。连接主义学派认为神经元不仅是大脑神经系统的基本单元，更是行为反应的基本单元。思维过程是神经元的连接活动过程，是通过大量突触相互动态联系着的众多神经元协同作用来完成的。

基于以上的思路，连接主义学派通过人工构建神经网络的方式来模拟人类智能——以工程技术手段模拟人脑神经系统的结构和功能为特征，通过大量的非线性并行处理器来模拟人脑中众多的神经元，用处理器的复杂连接关系来模拟人脑中众多神经元之间的突触行为。显然，相较符号主义学派，连接主义学派更看重智能赖以实现的“硬件”。这种方法在一定程度上可能实现了人脑形象思维的功能，即实现了人的右脑形象抽象思维功能的模拟。

图1-8 弗兰克·罗森布拉特

连接主义学派最主要的成果是人工神经网络技术。早在1943年，生理学家沃伦·麦卡洛克（Warren Mc Culloch）和数理逻辑学家沃尔特·匹兹（Walter Pitts）就提出的形式化神经元模型。他们提出神经元形式化

的数学描述和网络的结构方法，为人工智能创造了一条用电子装置模仿人脑结构和功能的新途径。此后，神经网络被不断改进：美国心理学家弗兰克·罗森布拉特（Frank Rosenblatt）将反馈学习算法引入神经网络中，英国科学家杰夫瑞·辛顿（Geoffrey Hinton）则提出将神经网络由一层改进为多层，美国心理学家大卫·鲁梅尔哈特（David Rumelhart）等人提出了多层网络中的反向传播算法，使多层感知机的理论模型有所突破。

神经网络最重要的改进出现在世纪之交。2000年，两位俄罗斯科学家弗拉基米尔·万普尼克（Vladimir Naumovich Vapnik）和阿列克谢·切沃内基斯（Alexey Yakovlevich Chervonenkis）提出了统计学习理论，并进一步提出了支持向量机模型。虽然统计学习在各个领域中都得到了广泛应用，但连接主义学派依然面临着难以解决的问题：科学家们虽然会向大脑学习如何构造神经网络模型，却根本不清楚这些神经网络究竟是如何工作的。智能仍然躲在黑盒子里，深藏不露。

行为主义学派的出发点与符号主义学派和连接主义学派完全不同，他们认为人工智能源于由美国数学家诺伯特·维纳（Nor-bert Wiener）建立的全新学科——控制论。控制论把神经系统的工作原理与信息理论、控制理论、逻辑以及计算机联系起来，其研究重点落脚于模拟人在控制过程中的智能行为和作用，如对自寻优、自适应、自镇定、自组织和自学习等控制论系统的研究。正是上述研究播下了智能控制和智能机器人的种子，并在20世纪80年代催生了智能控制和智能机器人系统。

在智能方面，行为主义学派并没有把关注焦点放在人类身上，而是投向了昆虫。昆虫虽然比人类低级得多，但其智能水平仍令计算机难以望其项背。从个体角度而言，昆虫可以灵活地摆动自己的身体行走，还能够快速躲避捕食者的攻击；从群体角度而言，大量昆虫聚集在一起时能表现出非凡的群体智能，还能形成严密的社会性组织方式。从更长的时间尺度看，生物体对环境的适应还会迫使生物进化，从而实现从简单

到复杂、从低等到高等的跃迁。

行为主义学派的机械代表作首推美国麻省理工学院教授罗德尼·布鲁克斯（Rodney Brooks）设计的六足行走机器人，它被视为“控制论动物”，是一个基于感知-动作模式模拟昆虫行为的控制系统。它们看起来的智能事实上并不来源于自上而下的复杂设计，而是来源于自下而上的与环境的互动。这就是行为主义学派所倡导的理念。另一方面，行为主义学派的算法代表则是美国科学家约翰·霍兰（John Holland）提出的遗传算法和美国心理学家詹姆斯·肯尼迪（James Kennedy）提出的粒子群优化算法。遗传算法对进化中的自然选择现象进行了高度抽象，通过变异和选择实现目标函数的最优化；粒子群优化算法则通过模拟动物的群体行为解决最优化问题。

行为主义智能的终极形式是由彼时就职于洛斯阿拉莫斯国家实验室的克里斯托弗·兰顿（Christopher Langton）提出的人工生命。所谓的生命或者智能实际上是从微观单元的相互作用而产生的宏观属性，这些微观单元既然可以是蛋白质分子，为什么不能是二进制符号形成的代码段呢？人工生命的研究思路正是通过模拟的形式在计算机数码世界中产生类似现实世界的涌现。

图1-9 罗德尼·布鲁克斯

可现在看来，行为学派带来的问题似乎比提供的解决方法还多。究竟在什么情况下能够发生涌现？如何设计底层规则使得系统能够以我们希望的方式涌现？这些问题尚未出现让人满意的答案，高级的智能也完全没有如期待般自然涌现，甚至没有丝毫涌现的迹象。

技术的十字路口
——人工智能谁领风骚

人工智能三大学派从不同的角度理解、定义和构造智能，事实上，它们之间还存在着很多微妙的差异和联系。

符号主义学派认为计算机是处理思维符号的系统，致力于用数理逻辑方法利用计算机形式化地表达世界。尽管按照这种方式来工作的专家系统已经在表达科学思维的某些方面达到了人类专家的水平，但这并不能制造具有自我意识的“人工智能”系统。因为从根本上来说万能的逻辑推理体系是不可能存在的，要计算机或智能机器完全模拟人脑的活动几乎是不可能完成的工作。

认知神经学表明人脑并非以线性顺序进行思维，而是以复杂的并行操作来处理感觉信息，连接主义学派正是据此主张从神经生物学的角度来模拟动物或人的大脑及各种感觉器官的结构和功能，力图寻找一种可以描述自然神经系统的方法，建立神经生理学模型。但人脑是一个异常复杂的组织，目前对人脑结构和活动机制的了解只是冰山一角，要建立一个与人类大脑相近的神经网络目前看来还是天方夜谭。

与前两者不同，行为主义学派从生物进化学的角度来研究人类的智能，认为智能是生物体对外界复杂环境的动态适应，人工智能只有从复制动物的智能开始，才能最终复制人的智能。基于以上观点，他们放弃了对智能的抽象描述的计算引擎，而是通过来自环境世界的情景、感应器内的信号转换以及机器人和环境的相互作用完成智能的构建。但是这一基于行为主义的感知——动作模式只能获得特定目标的行为，而在意向性、创造性方面还有难以克服的困难。

三大学派分别从高、中、低三个层次来模拟智能，但现实中的智能系统显然没有这么简单。如何将三大学派的观点融会贯通，将会是人工智能的下一个突破口。

第二章　安能辨我人或机
——通用人工智能理论

1970年，美国图兰大学的心理学家小戈登·盖洛普（Gordon Gallup Jr.)进行了著名的镜子实验（The Mirror Test），这一实验的用意在于测试动物的自我意识——是否能像人一样在镜子中分辨自己。盖洛普在黑猩猩的脸上画了小红点并观察它们如何照镜子。实验结果表明，和人的经验一样，照了镜子的黑猩猩立刻意识到“自己”的脸上出现了奇怪的红点，抓耳挠腮想要抹掉这些不速之客。而当实验对象换成猕猴——一种比黑猩猩低等不少的灵长类动物，结果就变得大相径庭：哪怕是照上一个月的镜子，猕猴也完全不会意识到镜子里就是它们“自己”。它们忙着每天和镜子里的“新朋友”打闹玩耍，根本无暇理会那些奇怪的小红点究竟是什么鬼。

1977年，美国人比利·米利根（Billy Milligan）因持械抢劫和校园强奸被起诉。但在法庭上，米利根的辩护律师证明了自己的代理人是多重人格症患者，作案时的比利事实上是另外的两个“自我”控制了他的身体和行为——事实上，米利根身体中潜藏的人格多达24个！其中既有南斯拉夫的共产党员，也有虔诚的犹太教信徒；既有打闷棍套白狼的小古惑仔，也有幼儿园年纪的英国小女孩。这番辩护成功说服了陪审团，米利根被无罪释放并进入精神病院治疗，他成为历史上第一个因为多重人格而免罪的人。米利根的不幸经历因纪实文学《24个比利》和《比利战争》而广为人知，根据他的经历改编的电影《分裂（Split）》也于2017年上映。在“我是谁”这个看似再简单不过的问题上，米利根的经历给出

了另外一种回答，自反映意识能力在内的那种元思维之“心”看起来并非理所当然。那么问题来了，机器是否能够具备类似的自我意识呢？

机器能思考吗？
——智能的本质在哪里

从前文的探讨中可以看出，对人工智能的思考与评价在很大程度上依赖于对“智能”的定义。因此，在讨论人工智能的前景之前，有必要对人类智能的本质进行一些阐释。但思维与智能这个问题本身恐怕用千页篇幅也不能尽述，在此只做挂一漏万的解读。

人类智能的本质是什么？这是认知科学的基本任务，也是基础科学面临的四大难题中最难解决的一个。根据自底向上的分析方法，人类智能的本质在很大程度上取决于“什么是认知基本单元”。目前的理论和实验结果表明，要分析认知基本单元是什么，合理的方法并非物理的推理或数学的分析，而是设计合理的认知科学实验。已有大量实验结果显示，从被认知的客体角度来看，认知基本单元是知觉组织形成的“知觉物体”。例如当人的视觉系统注意一只飞鸟的时候，它所注意的是整只鸟（知觉物体），而不是鸟的某个特性（形状、大小、位置等），尽管飞行时鸟的特性在时刻变化，但它作为同一个知觉物体的整体性却始终保持不变。用学院腔的话说，知觉物体概念的直觉定义正是在形状、位置等特征性质改变下保持不变的同一性。

知觉物体概念的形成具备其特殊的物理基础。脑神经科学研究表明，人脑由大约千亿个神经细胞及亿亿个神经突触组成，这些神经细胞及其突触构成一个庞大的生物神经网络。每个神经细胞通过突触与其他神经细胞进行连接与信息传递。当通过突触所接收到的信号强度超过某个阈值时，神经细胞便会进入激活状态，并通过突触向上层神经细胞发送激活信号。人类所有与意识及智能有关的活动，都是通过特定区域神

经细胞间的相互激活与协同工作而实现的。

作为一个复杂的多级系统，大脑思维功能源于功能的逐级整合：各神经元的功能被整合为神经网络的功能，各神经网络的功能被整合为神经回路的功能，各神经回路的功能最终被整合为大脑的思维功能。巧妙的是，在逐级整合的过程中，每一个层次上实现的都是“1+1＞2”的效果，在较高层次上产生了较低层次的每个子系统都不具备的“突生性质或功能”。这就意味着思维问题不能用还原论的方法来解决，即不能靠发现单个细胞的结构和物质分子来解决，揭示出能把大量神经元组装成一个功能系统的设计原理，才是问题的实质所在。

大脑利用定型的电信号处理它接受和分析的所有信息，外部世界的种种刺激都被量化为或弱或强的生物电流。有充分证据表明，感觉神经元仅对其敏感的事物属性作出反应。从信息科学的角度理解，这意味着感觉神经信号就是神经元对其敏感属性的编码。外部事物属性一般通过光波、声波、电波等模拟物理信息作为输入刺激人类的生物传感器，而感觉神经元输出的感觉编码是一种可符号化的心理信息。因此，感觉属性检测是一种将数值信息转化为符号信息的定性操作过程。更直白地说，感觉神经元实质上是将其敏感的事物属性从包含它们的物理刺激中抽取出来，并转化为该属性感觉映象的定性检测器。

感觉将事物属性转化为其感觉编码，不仅让大脑意识到该事物具有其检测的属性，还在事物属性集与人脑感觉记忆集之间建立起对应关系，所以感觉属性检测又叫感觉定性映射。如果说大脑是靠逐级整合各级神经网络的功能才形成其思维功能的话，那么由于感觉神经元的输出是各种简单属性的感觉映象，其高层神经网络整合的对象就只能是各种简单属性的感觉映象。于是，大脑怎样从这些简单属性的感觉映象中将对象的心理表象或记忆模式整合出来，并利用它们进行各种思维操作，就成了思维与智能研究中的关键问题。

神经网络整合事物各简单属性的感觉映象，得到的是该事物的整合

属性的感觉映象。比如大脑整合苹果的颜色属性（如红色）和形状属性（如圆形）的感觉映象的结果，应得到该苹果又红又圆这个整合属性的感觉映象。反过来，事物某整合属性的感觉映象又应该是该整合属性的各个因子属性的感觉映象的整合，苹果又红又圆这个整合属性的感觉映象，应该是红和圆这两个因子属性的感觉映象的整合。因此，在感觉映射下，事物属性结构与其感觉映象结构之间应保持不变，也就是说，感觉映射应该是事物属性集与其感觉记忆集之间的一个同态映射。通常所谓人脑认知结构是外部世界（结构）的反映，只是感觉同态的一种通俗说法而已。反过来，若感觉映射的确是一同映射的话，那么，事物属性的感觉映象结构与该事物的属性结构之间就应该是一致的。

根据感觉同态原理，事物性质（或质的定性）由此而产生的模糊性、非单调性和矛盾性等各种不确定性，不仅要同态地映射到人的感知记忆集中，而且，要在人的各种思维活动中反映出来。也就是说，所谓人类思维与智能的各种不确定性，实质上只是事物性质（或质的定性）的各种不确定性的表现而已。反过来，人类自身感觉阈限，或定性基准也要产生各种变化，感觉输出的各种不确定性感觉映象，又被整合为更高级的、带有各种不确定性的整合属性、关系和结构的记忆模式，人们利用这些记忆模式作其思维的素材，其思维当然会产生各种不确定性。根据感觉同态原理，思维中的这些不确定性是可转化到相应的事物属性集中来加以同态地讨论的，因而，也是可由属性坐标系加以数学表达的。

如果要对以上枯燥难懂的文字加以形象的梳理，就是这样一幅图景：人类自从他能被叫做人的那一天起就具备识别物体的能力了——这是剑齿虎，那是长毛象，手里的是棍子。其实进入我们眼睛的不过是不同波长不同数量的光子，是我们的视网膜和大脑的视觉皮层把这些光子进一步加工为不同的属性——这就是信息抽象的过程，作为加工工具的神经网络则部分来自于祖先的遗传，部分来自于自身的进化——最后在我们的脑中能够找到见过的动物的脑细胞，形成了对事物的抽象。在不

同的情境下，祖先会发出或表示危险或表示安全或表示高兴或表示悲伤的叫声，语言正是对大脑对不同叫声的抽象结果。有了语言，人类的交流就更加自如了。有时候要表达语言也说不清楚的意思时，就只好拿棍子在地上画画，图画最终抽象成了文字，使人类能够更全面更持久地传递各种经验。

大脑的抽象功能把人类和动物区分开来，帮助人类学会了耕种、取火、制作武器等技能，建立了文明，又抽象出文学、数学、物理、化学等分门别类的学科，这些学科反过来有帮助人类创造出今天的社会，科学知识只是人类知识的很小的部分，它只是自然规律信息的投影。人类进步的过程，也是创造抽象信息和使用抽象信息的能力逐步提高的过程。

只有以对人脑的物理性认识作为基础，探讨智能才是水到渠成。在自然界中，智能并非人类的专利，绝大部分动物、甚至某些植物都具备智能，只不过其水准远远低于人类的水平。但这种差异只是数量上的差异，究其本质，自然界的智能都体现为对信息的抽象，蚁群、蜂群、鱼群等群体性动物都表现出集群智能特征，其本质也无外乎是信息的抽象与共享。

归根到底，人脑进行的是复杂度超高的抽象计算，其智能化程度绝非现有计算机的水平可以比拟。眼睛、耳朵等感知器官与注意、意识等高级认知功能之间有高强度的交互作用，而不仅仅是实现信息获取这么简单的功能。人工智能的目标并不是模拟出和人脑功能毫无二致的计算机——那还要人脑做什么呢？更重要的恐怕还是实现互补的功能。计算机和人脑对信息的表征有着本质的区别，实现计算的架构也完全不同，能够获取的样本数也有差别。从第一台计算机诞生到现在不过区区七十年的历史，可人类却是经历了千万年的进化才达到今天的水平，因此要求计算机算法具有与人的智能类似的准确率和推广能力也无异于水中捞月、雾里摘花。

但问题恰恰在于计算机服务的对象是人，实际需求也是辅助人来实现类似的认知功能，用户不可避免地将计算结果与人的认知过程作比较，并用人的处理结果来评价计算机算法的优劣。不过，估计用户不会满足于一个计算机识别系统只能正确地识别一类物体，他们会很自然地要求设计的系统能够像人一样处理视听觉信息，这正是推动计算机像人一样工作的动力。

熟悉的麻木
——人类智能有多强大

世界顶级的计算机科学家高德纳（Donald Knuth）曾经如此评价人工智能：“人工智能已经在几乎所有需要思考的领域超过了人类，但是在那些人类和其他动物不需要思考就能完成的事情上，还差得很远。”人类的心智活动运行得如此流畅，以至于我们把它当作理所当然，却对它的精巧与美妙浑然不觉。直到运用科学与技术尝试解释其运作原理时，其复杂精密的设计才让我们后知后觉地赞叹不已。正是人工智能的发展在一次次地提醒我们：人类能进化到今天的样子是多么伟大的成就。

人类的大脑就像是一台杂乱地拼装在一起的器件，虽然低效、笨拙，兼之深奥难解，却还能正常工作。无论从哪个层面看，大脑都是个设计拙劣、效率低下的团块，可又出人意料地运作良好。大脑不是一台快速且万能的超级计算机，它不是一个天才在白纸上即兴完成的杰作。大脑是一座独一无二的大厦，沉淀着数百万年的进化历史。大脑很久以前对某个问题形成了特定的解决方法，经年累月一直使用它，或者加以改进用于其他用途，或者严格限制其改变。用分子学家弗朗索瓦·雅各布（Francois Jacob）的话说来就是：进化是个修补匠，而不是工程师。可正是修补匠的缝缝补补，补出了宇宙中最令人叹为观止的智能奇迹。

对形状的判断与分类是人类的基本技能，这一点在文字的处理上尤为明显。任何一个计算机操作系统中都有字体册，里面存储着百余种字体，每一种都代表着文字或字母不同的显示方式。但无论字体如何变化，一个识字的人都不会把汉字“土”认成“士”，也不会把英文字母“i”认

成“j”——这背后隐藏的正是惊人的抽象能力，只不过因为人类已经掌握而显得稀松平常。与此形成鲜明对比的是，即使计算机的形状辨认能力在今天已经得到长足进步，却依然难望人类的项背。

对这一事实最有力的说明就是验证码的应用：几乎所有的网站都要求你在注册时辨认并输入一串扭曲的字符，其目的在于证明服务器另一端的你是人类而非机器。这串简单的字符却有着冗长的学名：全自动区分计算机和人类的图灵测试，英文缩写为CAPTCHA——也就是通常所说的验证码。验证码的出现恰是人类智能绝妙的体现：识字的小朋友都能完成的任务，迄今为止却没有任何计算机算法能够做到。

在验证码中，字母和数字被嵌入杂乱无章的场景之下，灰度也被仔细调整，人为加入的各种噪声则让画面看起来更随机；为了防止根据频域特征识别，场景中还会加入线条等元素以破坏图像的大尺度特征。虽然从统计特性上看，噪声和目标字符没有差别，但是这些元素组合在一起显然导致了质变。这质变对人而言无关痛痒——在这类复杂的场景中正确地分割和识别出物体是小菜一碟。虽然场景中的颜色、形状、朝向千变万化，各不相同，互相遮挡程度很深，甚至有些物体的背景都在运动，但这些特点丝毫不会影响识别的准确性，视觉功能正常的人做这类任务的准确率几乎是100%，这突出地体现出人类视觉和计算机视觉的之间的差异。这一技术目前被互联网网站广泛应用，着实是对计算机视觉无声的嘲讽。

图2-1 典型的验证码图片

隐含在验证码背后的是关于人类认知的迷人话题：当物体的视觉信息经过视网膜和外膝体之后，会在注意环路的调节和控制下被初级和高级视觉皮层完成逐级的表征和加工过程，从而引导大脑发现物体自身醒目的特征和需要提前注意的特征。如果对面是一张人脸，在感知、区分和识别的过程中，我们会先后实现如下的判断“是人脸还是桌子？”“这个人我认识吗？他叫什么？在哪工作？”等过程，之后，记忆系统还会进一步完成一系列的高级加工：再认和回忆——某时某刻我在某地见过他、语义判断——他的名字有什么特殊的含义、以及情绪加工——这个人看起来绝非善类，我不喜欢他。

所有这些认知都可以归结为一个高度抽象化的加工模型。在这个模型中，信息的加工具有从简单到复杂的层次化特征，在每个层次上都有相应的表征，提取从线段朝向、简单特征组合到复杂特征等不同的特征，感知、识别等认知加工也是由这些不同表征的组合完成的。表征和加工的物质基础是神经元，大量神经元构成的群体的同步活动是实现表征和加工的生理学机制。单个神经元只能表征极为简单的信息，但当它们通过神经电活动有节律的同步震荡整合在一起时，复杂的功能就诞生了。从信息科学的角度看，整个加工过程的实现可以理解为复杂的多次特征提取过程，提取的特征从简单到复杂，多次组合，甚至“概念”这种

十分抽象的特征也可以被提取。

但如果人类的认知过程只是提取当前信息的特征并进行分类这么简单的话，它可不值得如此大费笔墨——认知还和注意、情绪等系统有着极强的交互作用，这些功能和认知密切相关。人的情绪对认知的影响绝非中晚期才启动的高级过程，它的作用远比我们想象的多。焦虑症、抑郁症等情感疾病的患者与正常人相比，对负性情绪信息有注意偏向，对带有负面色彩的情绪刺激更容易关注，这种注意偏向发生在视觉感知的早期阶段，其机理至今还笼罩在迷雾之中。

从生理角度上讲，认知过程中不同的加工是由大脑的不同区域实现的，这些区域并非各自为战，大部分情况下会协同工作。这种类似多重备份的机制的优势在于大脑的部分区域出现功能丧失时，不会导致识别功能完全瘫痪。在神经外科中经常出现令人称奇的病例：一位小脑完全缺失的患者竟然可以长期存活，她的不正常表现仅仅是走路不稳和轻微的发音不清；另一位患者大脑中连接两个最重要的语言区的弓状束纤维受到肿瘤压迫，这让她无法说出呈现在眼前的钥匙的名称，但是能明确描述出它是用来开门的，还亲手操作用法，情急脱口而出说这是瓶起子。某些脑区的缺失不但不会影响存活，甚至对正常生活的影响也远非我们想象中那般严重，这不得不说是进化的奇迹。

我们作为人类足以引以为傲的还不止于此。我们能够自由地控制四肢和身体，这同样是工程学上的卓越成就。人类的双手可称得上终极进化版瑞士军刀，在大脑的精确控制下能够执行几十种相当于卡盘、握把、夹具和钩子的功能。双腿亦是如此：这是真正意义上的全路况交通工具，它帮助我们征服了地球上从陡峭的山坡到崎岖的深谷的各个角落。站立、行走、奔跑、跳跃——在大脑的支配下双腿可以完成无比复杂的动作。目前，售价几千元的扫地机器人既爬不出地毯的边缘，也越不过最矮的门槛。要知道这些机器人是以轮子来驱动的，对于包含人工关节的机器义腿的控制比对轮子的控制还要复杂百倍。

相比于对身体的控制，更难以解释、也更让人着迷的是思维的自主性。人类心目中的常识实际上是个极端复杂的推论系统，更是个意义深远的科学问题。人或动物某一时刻只能处在一个位置、人的身体与思维是一个整体、人死则不能复生……认识这些看似理所应当的事实对人类而言荒谬可笑，可最强大的计算机也未必知道这些常识：他们没有日常生活中的体验、没有来自其他人的教导、也没有关于事实基本构成的核心概念。

人类智能与人工智能之间可以通过信息处理的桥梁联系起来，两者都是通过对信息的处理来组织事物。人类智能的运行模式可视为一个信息系统：来自外界的输入信息会刺激大脑产生反应，大脑根据逻辑等先验规则对信息进行处理，处理得到的输出再反馈给外界。但相比于基于计算机的人工智能，人类智能还有无数奥秘有待破解，这也给人工智能的发展路径笼罩上层层迷雾。

学习、归纳与推理
——这才是人工智能

创造出像你我一样具有自我意识和思考的人工智能无疑是世界上最具挑战性的问题之一，新的存在总是想窥探造物主的秘密。事实上，这也是达特茅斯会议的初心所在。可眼下，主流人工智能界已变得越发谨慎，目标也开始转移到针对某类特殊问题、特殊功能、特殊领域设计算法问题求解。在坚持传统的研究者看来，人工智能早已进入“脑死亡”的状态。

单纯的信息提取对计算机而言可能只是小菜一碟，但人类经过了千万年的进化，要复制这久经考验而形成的智能水准现在看来还是路漫漫其修远兮。计算机和人脑的运行规则显然是两个相差甚远的范式。目的在于提取信息的人类智能本就不求在精神世界里客观准确地再现物理世界，其终极目标是“生存”：用最合理的代价，获取最大的生存优势。

正因如此，物理世界中的不同信息在精神世界中具有不同的权重也就理所当然。体积相对较小的手指、舌头等重点区域占据着感觉运动中枢里的大部分皮层，在视觉上也只有对应中央视野的视网膜具有较高的空间分辨率和颜色分辨率，而更广泛的外周视野只对外界物体的突变敏感。人类正是通过知觉组织的选择性注意机制，直接感知输入信号中的大尺度不变性质，而忽略大量的局部的小尺度特征，其生态意义就在于对知觉对象进行信息抽提。

选择性注意机制的生理学基础是模块化的层次结构和分布式表征。功能模块化假设认为，大脑是由结构和功能相对独立、专司特定认知功能的多个脑区组成。这些模块组成复杂的层次结构，通过层次间的传递

和反馈实现对输入信号的主动调节。大量脑成像的研究实验也支持了这一假设，特别是视觉研究发现了非常详细而复杂的功能模块及其层次结构。另一方面，分布式表征的假说认为，认知功能的神经机制是相对大范围的分布式脑状态，而不是特定脑区的激活与否。

当然，人脑不仅仅是个针对刺激产生反应的简单系统，其内生性活动的强度甚至高于反应性活动，人脑在所谓的静息状态下的耗氧量与任务状态下相比差别很小这一事实也验证了以上观点。遗憾的是，由于内生性活动无论是定性观察还是定量观察都非常困难，导致其被长期排除在认知科学的研究主流之外。随着脑成像技术的发展，功能连接成为分析静息态大脑自发活动的有力工具。特别是默认网络的发现，创立了强调内生性活动的全新脑功能成像研究范式。默认网络被认为涉及警觉状态、自我意识、注意调控以及学习记忆等心理认知过程，已被广泛应用于社会认知、自我、注意、学习、发育、衰老机制的研究，有力推动了各种脑生物指标的完善和脑疾病的治疗，这些疾病包括阿尔兹海默病、帕金森病、抑郁症、精神分裂症和自闭症等等。

相比之下，计算机在很多事情上都显示出其“笨拙”。在安静的环境中，语音识别的准确率已经可以达到90%，但是在存在方言、噪音、远场等问题的实际环境下，其准确率就会不忍直视。最常见的例子就是鸡尾酒问题。在酒会上，即使周围十分嘈杂，有许多人同时交谈，人仍然能够顺利地分辨出和我交谈的那个人的声音，机器的语音识别算法在这样的场景下则基本等同于聋子。

如何从狭窄的特定领域智能迈向更通用的智能？这里的“通用”意味着机器能在没有编码特定领域知识的情况下解决不同种类的问题。作为人类，我们希望机器能够像我们一样进行判断与决策，而非限于在规则的框架内给出解释性的回答。通用智能应当具备处理多种类型的任务和适应未曾预料的情形的能力，例如，它无疑可以实现“正义”和“公平”这样的概念：事实上，人工智能对法律系统的影响已经近在眼前。

当莱特兄弟和其他先驱者停止模仿鸟类并开始学习空气动力学时，对人工飞行的追求才开始走向成功，对通用智能的追求过程同样如此。人工智能不需要将重点放到模仿大脑的生物过程上，而应该尝试理解大脑所处理的问题。可以合理地估计，人类使用了任意数量的技术进行学习，而不管学习在生物学层面上的实现方式。对通用人工智能来说也是一样：它将使用穷举搜索树，它将使用基于规则的系统，它也将使用模式匹配。但不管使用什么，它在处事方式上应该像个人类，而非仅仅说出一些似是而非的“人话”。

通用人工智能（Artificial General Intelligence）由德国计算机科学家马库斯·胡特（Marcus Hutter）提出，他对这一概念给出了精当的定义：

智能是主体在各种各样的纷繁复杂的环境中实现目标的能力。

把这一概念翻译成计算机能够理解的语言，得到的结果就是通用归纳模型和序贯决策理论的结合。通用归纳将归纳转化为预测，而预测的关键是对数据的建模或编码表示，预测的精度依赖于对模式的掌握程度。诸如分类、类比、联想、泛化等与智能相关的要素，都可以被广义地理解为对模式的追求，对计算机而言则是找寻最优编码。但找寻最优编码的过程无法通过解析方式执行，只能通过试错不断逼近。

试错的实质就是序贯决策理论，它研究的是在客观概率分布已知但具体状态不确定的动态环境中，主体如何寻求最大化期望效用。它从初始状态开始，每个时刻根据所观察到的状态和以前状态的记录，依照已知的概率分布，从一组可行方案中选用一个能够获得最大化期望效用的最优方案，接着观察下一步实际出现的状态，然后再作出新的最优决策，如此反复进行。但最关键的问题是，如果这种客观的概率分布未知怎么办？

建立了未知概率分布下的决策体系，就建立了通用人工智能，这也正是“算法概率”的宗旨与要义。算法概率的含义比较复杂，对这一概念

的探讨已超出了本书的范围。但简而言之，其核心是用可控的主观概率代替未知的客观概率，不同的主观概率则来源于对描述客观世界的不同算法的反向推演。如此一来，归纳是一个不断试错的学习过程，算法概率使得我们可以根据经验不断修正信念、逼近真实的客观概率，再借助序贯决策理论帮助我们追逐效用，能够自动适应各种可能环境的超级智能体就横空出世了。

通用智能在指明人工智能发展方向的同时，也给人工智能的水平设置了上界。如果现实世界没有那么纷繁复杂——或许现实世界真的没有想象中那么纷繁复杂，一个简单的幂律分布就能在各种尺度上各种环境中支配着各种看上去不相干的现象；或着某些复杂的模式可能是存在的，但某些可计算但异常复杂的模式可能不过是我们数学上的抽象构造，未必真的都会被物理例示——那么，算法概率就可以迅速地收敛到真实的现实世界，某个可计算且足够强大能适应足够复杂的环境的智能体也可能不难找到；但如果现实世界确实包含高复杂度的各种可能的模式，那么，简单的数学理论就无能为力了。

长久以来，通用智能概念的创始人胡特如一个特立的苦行僧，在熙熙攘攘的人工智能领域踽踽独行。在这个各路算法大行其道的狂欢时代，对终极问题的思辨与追寻反而显得格格不入。

不完备性定理
——哥德尔的“诅咒”

谈论通用人工智能，哥德尔不完备性定理就是个不可回避的问题。这一定理由奥地利数学家库哥德尔提出——就是气走冯诺伊曼那位老兄，是对希尔伯特提出的23大难题中第二个的回应。遗憾的是，哥德尔发现的不完备性定理对第二问题给出了否定性的答案，颠覆了数理逻辑大厦的基础，一度大大动摇了人们对于公理化方法和数学，甚至于整个科学的信心。当提问者希尔伯特得知哥德尔的研究结果后，哀叹“哥廷根数学已死”；撰写《数学原理》的罗素得知哥德尔的研究结果后，心灰意冷转而研究哲学，并于1950年以另一部巨著《西方哲学史》摘取诺贝尔文学奖，真可谓塞翁失马焉知非福。

虽然意义重大，但不完备性定理的表述却令人惊讶地简洁，第一不完备性定理内容为：

在任何包含初等数论的形式系统中，都必定存在一个不可判定命题。

为了通俗地解释不完备性定理，我们先来看一句话：

这句话是错的。

这句话到底是对还是错呢？如果它是对的，得到的结论就和判断本身相悖；如果它是错的，又证明它代表的实际判断是对的。无论如何，这句话带来的都是自相矛盾的结果。这种逻辑上自相矛盾的论断就是悖论。由于这句话中涉及了对自身的指代，这类悖论也因而被称为自指悖论。自指悖论的出现挑战了非黑即白的世界观，也为不完备性定理的证

实埋下了草绳灰线。

不完备性定理针对的是数学的公理化系统。在希尔伯特眼里，公理化系统应该同时具备两种优雅的特性：一致性和完备性。一致性指公理化系统中不存在矛盾；完备性指所有真命题都可以由公理化系统证实。如此一来，公理化系统，乃至整个数学就成为一个自洽的整体，要想获得真知，只要在这个超级的公理化系统中不停地推导就可以了。遗憾的是，希尔伯特理想中坚不可摧的数学城堡终究只是不堪一击的盐沙之基，它的命门就在于对自指的迷惑：

本数学定理不可以被证明。

这条自指语句与前文的悖论有异曲同工之妙。首先，这个数学命题所讨论的对象不是别的，恰恰是它自己。“本数学命题”就是对整个命题的指代。其次，该命题给出了一个逻辑判断，即这条命题是不可以被证明的。这个句子本身似乎并没有那么邪恶，然而只要我们一开始用逻辑的头脑解读它，它就变成了一句魔咒，直接摧毁了希尔伯特的完备一致性猜想。

图2-2 库尔特·哥德尔及其签名

根据逻辑排中律，这条数学命题非真即假。如果这个数学命题是真命题，并且根据它自己的论述，它不能被证明。于是，我们得到了一条真理，但却不能被我们的数学公理化系统所证明，因此，希尔伯特要求的完备性不能得到保证。如果这个数学命题是假命题，这意味着“本数学命题不可以被证明”这个命题是可以被证明的。于是，从公理出发，我们能够得到“本数学命题不可以被证明”这一命题。而按照假定，“本数学命题可以被证明”是真理，所以根据完备性，它也必然是系统中的定理。于是，正命题和反命题同时都是系统中的定理，一致性遭到了破坏。

综上所述，我们可以断言：对于一个具备自指能力的数学公理化系统，一致性和完备性不能兼得。这便是第二不完备性定理。当然，在哥德尔的原始论文中，所有的表述都是使用严格的数学语言来表达的。

不完备性定理的应用不仅仅限于纯数学领域，它对一切抽象性问题的分析都产生了深远的影响，自然也包括了计算机科学和人工智能的发展。人类之所以会把实现人工智能的期望寄托在计算机身上，其基础在于“认知的本质是计算”这一著名论断。截至目前，所有的计算机都未能超出图灵机的范畴，也就必须遵循数理逻辑定义的规则。从“认知即计算”的角度出发，基于计算机的人工智能如果想要达到近似人类的思维能力，也必须建立起“自我”的概念，这无疑会导致自指的出现，也将成为不完备性定理的活靶子：如果计算机能在运算中制造出一个代表自身的符号，那么哥德尔制造悖论的方式就可以在计算机中造出不可证实也不可证伪的飘渺仙境。据此，基于图灵机理论模型的计算机绝无可能拥有代表自我的符号，也就绝无可能达到人类智能的水平。毕竟人类既能执行逻辑思维又能超越逻辑思维（谈恋爱时讲逻辑？那你注定要孤独一生哟！），而直到今天，一切试图以纯逻辑来描述世界的尝试都以失败告终。

利用不完备性定理预测人工智能的前景并非民间科学家的自说自话，主流学界也早已发声表态。达特茅斯会议召开后仅仅5年，美国哲学家约翰·卢卡斯（John Lucas）便发表论文《心、机器、哥德尔》，提出了著名的“卢卡斯论证”，以激烈的言辞试图用哥德尔定理证明计算机的智能水平无法达到人类水准：“依我看，哥德尔定理证明了机械论是错误的，也就是说，心不能解释成机器。”因为，“无论我们造出多么复杂的机器，只要它是机器，就将对应于一个形式系统，接着就能找到一个在该系统内不可证的公式而使之受到哥德尔构造不可判定命题那种程序的打击，机器不能把这个公式作为定理推导出来，但是人心却能看出它是真的。因此这台机器不是心的一个恰当模型。人们总想制造心的一种机械模型，即从本质上是‘死’的模型，而心是‘活’的，它总能比任何形式的、僵死的系统干得好。”

1989年，卢卡斯的牛津同侪，英国数学家和物理学家罗杰·彭罗斯（Roger Penrose）在风靡全球的著作《皇帝新脑——计算机、心智和物

理定律》中发展了卢卡斯论证，以大量笔墨试图基于不完备性定理推导出“人心超过计算机”的结论，被称为“对哥德尔定理令人吃惊的强应用”，在学界引发了广泛争议。彭罗斯的一个强硬论证是：根据哥德尔定理可以“像在数学中所做的那样，严格证明”数学真理的概念不可能包容于形式主义的框架之中，数学真理是某种超越纯粹形式主义的东西，人类判断数学真理的过程是超越任何算法的。这是因为，意识是我们赖以理解数学真理的关键，这种意识使我们能够借直觉的洞察力“看出”某些在数学形式系统中不能证明的数学命题的真理性，而这种意识是不能被形式化的，它必定是非算法的。因此人工智能绝不可能超越人类心智，所谓强人工智能不过是专家所钟爱的，与皇帝的新衣一般无二的一副“皇帝新脑”而已。

计算机是人类为了自身目的而设计制造的，这种制造者与被制造者之间的强关系将人置于一个面对面地统治机器的绝对优越地位，这种地位究竟是一种社会学意义上的优越，还是计算机和人的智能的本质特性所决定的？是像哥德尔断定的，存在与物质相分离的心能超越任何计算机去发现和证明某些数学定理；抑或如彭罗斯断言，人心具有一种特殊的能力，这种能力是建立在迄今未予发现的某些物理学规律的基础上，而且人心能超越任何计算机实现非算法的运算？这些都是我们需要深入探讨的问题。

人工智能，甚至整个认知理论正在面临着一场研究纲领的变革。在哥德尔不完备性定理的阴影下，基于图灵可计算概念的“认知可计算主义”研究纲领已经显示出其极大的局限。如果以“认知的算法不可完全性”为核心的研究纲领取而代之，人类就必将探索新的非图灵机概念来尝试解决人工智能更深层的问题，以摆脱在理论和实践上的困境。当然，解决这些问题除了靠哲学思辩，更需要依赖于科学的进展和精细的逻辑应用的研究。解决人类智能的极限和人工智能的极限问题，除了与绕不开的哥德尔不完备性定理有关外，还需要对大脑和计算机更精细的模型做更大胆的研究，而且还需要将学习、问题求解、对策理论与实数

论、逼近论、概率论和几何学知识结合在一起，探索其如何对问题的解起实质性作用。

第三章　从深蓝到阿尔法狗
——人工智能的技术演进

多年以后，面对韩国九段李世石（Lee Sedol）与阿尔法狗（Alpha Go）的厮杀，加里·卡斯帕罗夫（Garry Kasparov）将会回想起，与深蓝（Deep Blue）对弈的那个遥远的下午。彼时的对手是一个庞大的超级计算机，记下变化多端的开局，以固定的逻辑决策应对中局，穷极所有可能性筛选残局。人工智能新生伊始，许多概念尚且是空中楼阁，计算机能够战胜优秀的人类围棋选手的想法不啻于天方夜谭。

2017年开年，人工智能企业Deep Mind公司又搞了个大新闻：继阿尔法狗以压倒性的优势4:1战胜围棋世界冠军李世石后，大师（Master）——升级版的阿尔法狗——在网络上又实现了对各路围棋大师的60连胜！这一结果令人大跌眼镜：因为在20年前，虽然升级版的深蓝同样战胜了来自俄罗斯的国际象棋特级大师卡斯帕罗夫，可即使最乐观的人工智能研究者也不敢断言，未来有一天计算机能横扫人类的围棋精英们。

取胜之匙
——深蓝的“算”与阿尔法狗的“想”

棋类游戏的核心在于根据棋局判断下一手的最优下法，深蓝通过穷举的方法在国际象棋的棋局中解决了这个问题。在64格的国际象棋棋盘上，深蓝的运算能力决定了它能算出12手棋之后的局面下的最优解，而身为人类棋手执牛耳者的卡斯帕罗夫最多只能算出10手棋，这多出来的2手棋就会成为左右战局的关键因素。可在围棋棋盘上，可以落子的点数达到了361个——别说12手棋，就是6手棋的运算量都已经接近于天文数字！这使得计算机相对于人脑的运算优势变得微不足道，走出优于人类棋手的妙手的概率也微乎其微，这也是为什么计算机会在围棋领域被看衰。

深蓝的核心在于”算”：利用强大的计算资源来优化目标函数。深蓝本身就是一套专用于国际象棋的硬件，大部分逻辑规则是以特定的象棋芯片电路实现，辅以较少量负责调度与实现高阶功能的软件代码。其算法的核心则是暴力穷举：生成所有可能的下法，然后执行尽可能深的搜索，并不断对局面进行评估，尝试找出最佳下法。

图3-1 1997年超级计算机深蓝对阵国际象棋特级大师卡斯帕罗夫

在深蓝的象棋芯片上，国际象棋的走棋规则被以硬件电路的方式嵌入到逻辑门阵列之中，不同棋子处于不同位置时的分值由软件预先计算好后也会写入硬件。对下法的判断则源于国际象棋的固有逻辑。在国际象棋中，最核心的逻辑就是子力价值的对比：马或者象等效于三个兵；车等效于五个兵；后等效于九个兵；王的价值是无穷大，因为失去王就输了棋局。但在评价棋盘状态时，深蓝会考虑更多的局面细节：如果同一方的兵在象前面，它就会限制象的移动，导致象本身的价值降低；如果同一个兵可以通过捕获一个敌方兵来打开车的行进路线，这个兵就并不会严重损害车的价值。这类对棋局细节的刻画有助于深蓝对局面做出更准确的判断。

深蓝的软件来源于与硬件协同工作的的专门设计。软件部分负责调度最多32个象棋芯片并行搜索，以及对大范围规划的局面进行软件评估。软件中还包含了从数十万局棋中抽取出来的开局书，少子条件下的残局数据库，以及同时代的美国特级大师乔尔·本杰明（Joel Benjamin）针对卡斯帕罗夫行棋风格而对以上开局与残局下法的专门优化。因此，深蓝背后蕴藏着的是古往今来各路高手的象棋智慧，说卡斯帕罗夫是光

明顶上独战六大门派高手的张无忌，其实也不为过。

图3-2 2016年人工智能系统阿尔法狗对阵围棋世界冠军李世石

可是，用穷举的方式来下围棋呢？

围棋的棋盘状态远比国际象棋复杂，以穷举法进行最优落子策略的推演无异于痴人说梦。事实上，顶级的围棋棋手更多地依赖模糊的直觉来评判特定的棋盘状态的好坏。但理性的推演与感性的判断之间似乎存在着不可逾越的巨大鸿沟，对于计算机程序而言，依赖直觉是不可能的事情。因此并没有显而易见的方式来将国际象棋领域的成功复制到围棋上——直到阿尔法狗的横空出世。

图3-3 国际象棋计算机深蓝

图3-4 人工智能系统阿尔法狗

阿尔法狗的核心则在于“想”。与专用硬件深蓝不同，阿尔法狗是一套能够运行在通用硬件之上的纯软件程序。它汲取了人类棋手海量的棋谱数据，并依赖人工神经网络（Artificial Neural Network）和深度学习（Deep Learning）从这些数据中学会了预测人类棋手在任意的棋盘状态下走子的概率，模拟了以人类棋手的思维方式对棋局进行思考的过程。阿尔法狗算法的形成可以分为三个阶段：

第一阶段——拜师学艺：阿尔法狗根据彼此无关的盘面信息模仿专家棋手的走法，通过海量盘面数据训练出一个监督式策略网络，这个策略网络随后就能以超过50%的精度预测人类专家的落子；

第二阶段——左右互搏：阿尔法狗将过往训练迭代中的策略网络与当前的策略网络对弈，将对弈过程用于自我强化训练，对现有策略网络的改进使阿尔法狗对弈当时最强的开源围棋软件Pachi的胜率达到85%；

第三阶段——融会贯通：阿尔法狗在自我对弈中随机生成新的训练数据，用以训练局面价值网络。价值网络、策略网络和蒙特卡洛树搜索相融合，用于预测和评估棋局未来可能的发展方式。

拜师学艺完成后，阿尔法狗就可以中规中矩地下一盘棋了。在某种意义上，这是一种意识流的下法，胜负不在算法的考虑范围之内。而左右互搏的目的就是引入胜负：让策略网络和自身进行对弈，来获得一个给定的棋盘状态是否为胜利的概率估计，以此作为对棋盘状态的评估方式。最后，通过将评估方式和对下法的搜索进行融会贯通，选择那个给出最高棋盘状态评价的下法。不难看出，阿尔法狗并非从一个基于很多围棋细节知识的评价系统开始，而是让神经网络和机器学习扮演核心角色。它使用了两个各司其职的神经网络：策略网络和价值网络。策略网络的作用是选择下一步的走法，可以降低搜索的广度；价值网络的作用是评估盘面优劣，可以降低搜索的深度。通过连续不断做出微小改进的方式构建策略网络和价值网络，阿尔法狗就形成了类似于人类棋手所谓的关于不同棋盘状态的直觉的效果（当然也使用了搜索和优化的思想）。

策略网络本质上是个监督式学习（Supervised Learning）的过程，通过学习千万数量级的职业棋手棋谱来训练落子位置的预测模型。它有着专一的目标：完全不考虑输赢的概念，只关注预测对手落子的精确性。在2016年1月刊发在著名科学期刊《自然》的封面文章中，阿尔法狗预测对手落子位置的正确率是57%，这个数据在和李世石对弈时显然又得到了相当的提升。阿尔法狗的策略网络与类似的传统算法区别有二：一是左右互搏的引入：通过基础版本策略网络和进阶版本策略网络

之间的对弈，让基础网络快速习得高手的落子策略，形成一个比进阶更进阶的策略网络，这个新形成的策略网络又被用来进一步提高原始的进阶版本策略网络……经过两千万次“青出于蓝而胜于蓝”的循环修正后，策略网络才达到现在的水准；二是局面判断的设计：选择下一步的走法时，策略网络的备选并非棋盘上的所有361个点，而是通过卷积核先排除掉一些最优解出现概率较小的区域，再在剩余的区域中找出可能的最佳位置，这样就可以排除一些有意为之的干扰棋路对整体局势的影响。这种机制固然会降低落子预测的精确度，却能使计算速度得到大幅度的提升。

如果说策略网络关注的核心是“知彼”，价值网络关注的就是“知己”：在当前的局势下，我下在哪个位置能得到最大的胜算。对胜算的估计既与当前的局面有关，也与向下预测的步数有关：能够预测的步数越多，得到的结果就越精确，计算量也会越庞大。在围棋中，求解精确解显然是不可能的，因而价值网络只能求出近似解，通过卷积神经网络来计算出卷积核范围内的平均胜率，最终的走法则留给蒙特卡洛树搜索（Monte Carlo Tree Search）来处理。此外，价值网络的训练不是通过对现有棋谱的学习，而是让两个阿尔法狗互相对弈——两者实力的接近确保了棋局的胜负完全由落子决定，而非一些其他的先验因素。这让阿尔法狗快速地累积出正确的评价样本，也解决了评价机制的难题。价值网络和策略网络的结合让阿尔法狗“知己知彼”，其百战百胜自然也在情理之中。

但这并不意味着阿尔法狗无懈可击：在人机大战的第四局中，李世石的一招妙手（白78手）让阿尔法狗掉进了陷阱。阿尔法狗完全没有意识到这步神仙棋有什么作用，直到几个回合之后才如梦初醒，然而为时已晚。在此之后，阿尔法狗开始频频下出不可理喻的走法，直到投子认负。

李世石的妙手妙就妙在刺中了阿尔法狗的盲区：它并不认为棋会下

到这里。可能阿尔法狗认为这步棋并非最优甚至并非次优，可能在于自身对弈的过程中这样的棋路从未出现……种种原因让它没有在深度学习过程中习得这个走法。所以一旦出现这种局面，阿尔法狗开始不知如何是好，在盲目应对的过程中丧失主动。劣势后频出的昏招事实上也是蒙特卡洛树搜索的固有结果：自知败局已定的阿尔法狗只能通过这样的招数，寄望于凭借李世石的失误扭转局面。好在李世石没有上当，漂亮地赢下这一局，也赢下全世界的满堂彩。

虽然仍有改进的余地，但与深蓝的区别正是阿尔法狗的突破之处：早期的计算机就已经被用来搜索优化已有的函数的方式，深蓝的特点仅仅在于搜索的目标是优化尽管复杂但是形式大多数由已有的国际象棋知识表达的函数，其思想却与人工智能早期的多数程序并无二致。更令人诧异的是，在整个算法中，除了“获胜”这个概念，阿尔法狗对于围棋规则一无所知，更遑论定式等高级围棋的专门概念。尤其在第一阶段的训练中，完全基于简单的盘面信息就能够达到相当可观的预测效果。这也是阿尔法狗和深蓝的本质区别：同是战胜了棋类世界冠军，深蓝仍然是专注于国际象棋的、以暴力穷举为基础的特定用途人工智能；阿尔法狗是几乎没有特定领域知识的、基于机器学习的、高度通用的人工智能。这一区别决定了深蓝只是一个象征性的里程碑，而阿尔法狗则更具实用价值。

围棋似乎也并不是人工智能与人类对抗的终结。美国卡内基·梅隆大学的两位计算机科学家创造了会玩德州扑克（Texas Hold'em）的人工智能李贝特斯（Libratus），而李贝特斯则在2017年1月的匹兹堡大河赌场的无限制德州扑克游戏中大获全胜——以180万筹码的优势战胜了四位人类高手。虽然这一人机对战的关注度与影响力远不及围棋对弈，可它表现出的却是人工智能的另外一面。扑克游戏的声誉来自于艺术更甚于科学，对于计算机来说，这也是不同于各种棋类的独特挑战：非完美信息博弈需要某种人类的狡诈——例如欺骗对手并且能够察觉到对方在欺骗你——这通常被认为是计算机的阿喀琉斯之踵。在扑克游戏中执行

均衡战略的关键是打出最强和最有潜力的一手牌的同时还保持不可琢磨，看起来李贝特斯在这方面也是个好手。

人工智能与人类的下一次对抗在哪里？我们能做的，只有拭目以待。

最初的一步
——模式识别

通过计算机来实现人工智能，最初的路径是模式识别（Pat-tern Recognition）。模式识别的黄金时代出现在20世纪80年代，它强调的是如何让计算机程序去做一些看起来很“智能”的事情，就像是有个人躲在盒子里伪装成机器的样子。模式识别技术的主要作用在于发现、区分、检测或提取存在于我们周围世界中的模式，这依赖于怎么从观察数据中进行信息提取和表示，结合背景知识，最终得到新知识和概念的形式化内容。学习的结果是得到一个用于表示模式之间相互依赖的形式化知识，以此更好地理解与解释观察数据。当模式的概念被形式化后，它就可以被应用于相同领域未知的用例，包括新的信息，例如对一个新对象进行标识，且对于新用例的处理应当遵从应用于原来用例的相同的演绎过程。

具体来说，人们在观察事物或现象的时候，常常要寻找它与其他事物或现象的不同之处，并根据一定的目的把各个相似的但又不完全相同的事物或现象组成一类。字符识别就是这其中最典型的一个例子。在不同的字体中，数字“3”可以有不同的写法，但所有的写法都属于同一类别。更为重要的是，即使对于以前从未出现过的“3”的写法，识字的人凭借直觉和智慧也能够把它划分到“3”所属的这一类别之中，而不是错误地辨别为“8”或着“B”。“模式”的概念正是源于人脑的这种抽象思维能力：只要认识这个集合中的有限数量的事物或现象，就可以识别属于这个集合的任意多的事物或现象。为了强调从一些个别的事物或现象推断出事物或现象的总体性质，这些个别的事物或现象就被称作模式。

模式识别意在学习人类（或其他生物系统）在所处环境中发现、区别和找出特征从而标识出观察结果的本领，这属于认知科学的范畴，是生理学家、心理学家、生物学家和神经生理学家的工作范围；同时也专注于开发和评价模仿或辅助人类识别模式能力的系统，这是数学家、信息学专家和计算机科学家的用武之地。模式识别中工程的观点则是试图建立模拟生物识别能力的系统，这方面的研究已经取得了系统的成果，也给人工智能的发展打下了良好的理论基础。

早期的计算机模式识别研究将重点放在数学模型的建立上。1958年，供职于美国康奈尔航天实验室的心理学家罗森布拉特（神经网络的先驱者，参见第一章）提出了一种模拟人脑进行识别的简化数学模型——感知机，初步实现了通过给定类别的各个样本对识别系统进行训练，使系统在学习完毕后具有对其他未知类别的模式进行正确分类的能力。1974年，供职于美国普渡大学的华裔计算机科学家傅京孙出版了专著《句法模式识别及其应用》，系统梳理了模式识别在自然语言处理中的成果。就职于美国加州理工学院的科学家约翰·霍甫菲尔德（John Hopfield）则于1982年和1984年分别发表了两篇重要论文，深刻揭示出人工神经元网络所具有的联想存储和计算能力，进一步推动了模式识别的研究工作，从而形成了模式识别的人工神经元网络方法的新的学科方向，也将神经网络这一新的研究议题推到了聚光灯下。

模式识别的流程可以概括如下：首先要通过各种传感器把被研究对象的各种物理变量转换为计算机可以识别的数值或符号的集合，这个集合被称为模式空间，相应的数值或符号则被称为信号。对模式空间的必要处理——去除噪声的干扰、排除不相关的信号——是抽取有效识别信息的基础。在数据的识别中，模式空间中的信号经过特征量的提取和变换后，被映射到新的空间上，这个新的空间就是特征空间。与原始的模式空间不同的是，特征空间之中的元素是相互独立的，任意两两之间都不存在相关性，这显然构成了描述信号的一组基本元素，这个过程也可以被看作特征抽象的过程。模型匹配正是借助特征空间上的基本元素进

行的：通过对输入的对象进行同样的空间转换，模式识别系统会输出对象所属的类型或者是模型数据库中与对象最相似的模型编号。为了提升模式识别的精确性，往往需要加入一些预先设定的规则以对可能产生的错误进行修正，或通过引入限制条件大大缩小待识别模式在模型库中的搜索空间，以减少匹配计算量。

目前，模式识别技术最成功的实际应用，非光学字符识别（Optical Character Recognition）莫属。光学字符识别本质上是利用光学设备去捕获图像，并从中读取出出现的文字。未来的办公室中很可能出现这样的景象：只要使用手机等具备拍照功能的智能设备对会议白板进行拍照，系统便能自动识别出照片中的讨论内容，分检出相关人员的后续工作，并将待办事项自动存放到各自的电子日历中。正是光学字符识别的出现，使这样的场景成为可能。

光学字符识别技术的发展经历了超过半个世纪的摸爬滚打。1965年的纽约世界博览会上，IBM公司展出了其研发的光学字符识别产品——IBMl287.这款元老产品只能识别特定印刷字体的数字、英文字母及部分符号，在当时却已经是了不起的成就。IBM公司的研发成果在日本被迅速跟进：数年之后，日本东芝公司和NEC公司先后研制出手写体邮政编码识别的信函自动分拣系统，并广泛应用在邮政系统中，让信函的自动分拣率达到92%。1983年，东芝公司又发布了印刷体日文汉字的识别系统——OCRV595，其识别率高达99.5%。20世纪90年代后，平板扫描仪的应用与普及又成为光学字符识别技术走向下一个高潮的东风：谷歌公司雄心勃勃的数字图书馆计划正是由此而来。

光学字符识别中的技术难点在于字符的辨认与区分，其技术手段包括模式匹配识别法和特征提取识别法。模式匹配识别法是将数字图像中的字符与已有数据库中的标准字符相比较，以找到最相似的匹配，寻找的过程通常是以迭代方式进行的。特征提取识别法则是将每个字符分解为若干独立的字符特征，比如汉字笔划中的横、竖、撇、捺、点等等，

再将这些特征与待识别的字符进行匹配，根据不同的笔划凑出识别结果。

模式识别的核心意义在于分类——也就是所谓“模式”的区分，而每个模式的区别又由其与众不同的特征决定。可数据的种类千变万化，要发掘出隐藏在表象下的特征绝非易事。20世纪90年代，研究者开始意识到数据才是更有效地构建模式识别算法的方法——这正是机器学习的思想。事实也渐渐证明，机器在没有人类的干预下学习得到的结果远比各位专家纯手工设定的分类规则要好得多。这也催生了更多人工智能领域的先进技术。

大脑的人工模拟
——神经网络

前文介绍了人工智能发展过程中产生的三大学术流派，眼下风头正劲的正是其中的连接主义学派，阿尔法狗的横空出世也正是拜连接主义学派的神经网络所赐。

人工神经网络听起来玄之又玄，既然神经网络属于生理学和认知科学的范畴，那人工神经网络岂不是要人为合成一套神经系统？其实不然，人工神经网络只是一组数学模型，只不过这一数学模型被用于模拟人类神经系统的架构与功能，所以才被仿生地命名为人工神经网络。

关于神经网络的作用，在国外的问答网站Quora上有非常通俗的描述：如果你去买芒果，但又不知道什么样的芒果最好吃，最简单的方法就是每一个都亲口尝一尝，吃完就知道个头大、颜色深的比较好吃，再买的时候选这种就行了。要是把这个方法套用到计算机上，让计算机“尝”一遍所有芒果，它就能够总结出关于芒果好吃判断标准的一套规律。有了这套规律后，一旦把新芒果的特征输入计算机，计算机就能够根据已有规则判断出芒果的好坏，岂不美哉！

根据规则来判断芒果的好坏这个问题，属于模式识别的范畴，而人工神经网络正是解决模式识别问题的主流方法。

人的大脑是自然界中最强大的神经网络。打过麻将的读者可能会知道“审牌”的说法：久筑长城的牌坛老手不用看牌，单凭触摸麻将牌上的纹路就能知道这张牌到底是五条还是八万。当然，从没接触过麻将牌的小孩子无论如何也摸不出到底是哪一张，打牌较少的新手要达到较高的

审牌成功率也并非易事，只有长期摸牌打牌的人才具备这种牌桌上的高级技能。在这个现象中，我们可以一窥人工神经网络的工作原理。

要解释人工神经网络的原理，不妨把审牌的问题做个简化：分辨出一张牌到底是九条还是一饼。这两张牌的区别明显：九条的牌面只包含竖纹，而一饼的牌面只包含圆圈纹，每一种牌面的特征都是一种独特的模式。作为人类的我们可以通过手指的触觉识别不同的纹理，但计算机显然没有这么聪明——它只认识数学模型。既然如此，我们就可以把不同的纹理抽象化成一个平面直角坐标系，它的横轴代表竖纹的强度，纵轴代表圆圈纹的强度。这样一来，九条和一饼就分别落在两个坐标轴上。沿着这个坐标系的对角线画一条线，就可以轻松地把九条和一饼区分开来。再来一张牌，经过量化后落在这条线上面的就是一饼，下面的就是九条。

这条线起到的就是模式识别中分类器的作用，在人工神经网络中则对应神经元的概念。神经元的实质就是分类器，它把由所有输入信号构成的空间一分为二，两边的元素分别属于不同的类。最简单的分类器就是二维空间上的直线，这一直线扩展到三维空间上就是一个平面，扩展到四维空间上就是一个超平面……依此类推。这样一个线性函数虽然实现简单，但其功能也有限——只能用来区分九条与一饼，用来区分八条和二饼、七条和三饼的话，就可能会出现判断错误的情况。要是再把纹路结构更加复杂的一万到九万引入判定的话，这样的线性模型就不再适用了。

汉字的笔划有横有竖还有弧，所以万牌从牌面上讲，不是简单的单一纹路，而是横纹、竖纹和弧形纹的有机组合。这样的纹路扩展从数学上讲，实质上是将模式的特征复杂化，模式之间的区别精细化，一条直线就没有办法将不同模式精确地区分开来。一个直观的改进方法是将线性的神经元改造为非线性的神经元，非线性的曲线或者曲面显然比直线或平面具有更高的区分度。但不同地域的麻将规则不同，北方的麻将里

会有红中发财白板，四川的麻将里有东西南北风，江浙的麻将里有梅兰竹菊。要进一步区分出这些复杂的牌面，在平面上只画一条线就不够了，只有把平面划分成更多的区域才能实现精确的判定，这对应的就是多层神经网络。

通常的审牌过程正是多层神经网络的处理方式：审牌时我们会先摸出大致的纹路，判定这张牌到底是简单的条牌或饼牌，还是更加复杂的万牌或花牌。经过这一步的分类后，再来具体判定是四条还是五饼，是三万还是南风。多层神经网络的工作原理也是这样：下层神经元的输出是上层神经元的输入，不同层次的神经网络使用不同的神经元来分辨输入信号的不同特征，经过多层神经网络处理后得到的不同区域还可以进一步进行交、并、异或等逻辑运算。这样一来，多层神经网络就可以表示出更复杂的空间划分，得到更精确的判定效果，其代价则是更高的计算复杂度。

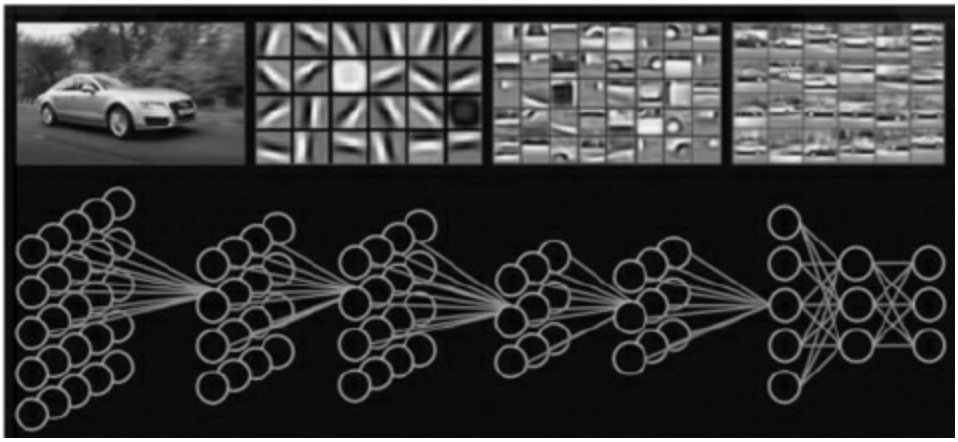

图3-5 神经网络识别图像示意图

如前文所示，通过与打麻将中审牌的对比，我们意在说明人工神经网络的作用是分类。实际的人工智能当然不能用于判断麻将牌，但在垃圾邮件识别，循证医学的临床路径、自然语言处理、图像识别等领域中都有广泛的应用。在这些应用中，人工神经网络都需要使用海量的分类器。审牌的能力可以通过长期的试错与反馈习得，但人工神经网络中的

大量参数又如何来确定呢？

说到这里就不得不涉及数学知识了。人工神经网络的数学本质是一种特殊的有向图，这个有向图可以由一层或多层节点组成，每一层的节点都通过有向弧指向上一层的节点，每一条有向弧都用一个权值来描述，同一层的节点之间则并无连接。输入层的节点按照有向弧的权值进行函数变换，变换后的输出传递给第二层的节点作为输入；第二层的节点如此这般执行同样的操作，其输出再作为第三层的输出。最后在输出层，哪个节点的数值最大，输入的信号就被划分在哪一类。

在此过程中，如何保证对输入信号的分类符合我们的要求呢？这就需要人为地对人工神经网络进行训练。所谓训练，就是通过负反馈的方式动态调整人工神经网络中权值的过程，目的就是使网络参数尽可能的与真实的模型逼近。既然我们希望网络的分类结果尽可能地接近真正情形，就可以通过比较网络当前的输出和真实值，再根据两者的差异情况来更新每一层的权重值来降低偏差。如果人工神经网络的预测值偏高，就调整权值使输出变低，反之则调整权值使输出变高。就这样不断调整，直到偏差小于某个特定的阈值为止，这时我们就认为人工神经网络达到了精确的分类。具体的训练方法则是反向传播算法：最开始输入层输入特征向量，网络层层计算获得输出，输出层发现输出和正确的类号不一样，这时它就让最后一层神经元进行参数调整，最后一层神经元不仅自己调整参数，还会勒令连接它的倒数第二层神经元调整，层层往回退着调整。经过调整的网络会在样本上继续测试，如果输出还是老分错，那就继续来一轮回退调整，直到网络输出满意为止。

人工神经网络的质量由三个要素决定：网络结构和节点函数，训练数据的质量和完备性、训练方法的合理性。其中网络结构类型和节点函数是预先设计的，训练数据和训练方法则是由外部导入的。网络结构和节点函数是决定人工神经网络质量的首要因素，运行机制符合实际事物的内在机理的网络才是高质量的人工神经网络。另一方面，如果训练数

据不典型、不充分、不完备，训练方法不适当，也会影响人工神经网络的输出精度。

计算机的无师自通
——深度学习

人工神经网络的本质是通过计算机算法来模仿、简化和抽象人脑的若干基本特性。起起落落之后，人工神经网络产业如今迎来了第三个高速发展时期，正是得益于深度学习的研究。

深度学习又被称为深度神经网络（Deep Neural Network），其基础也是人工神经网络，“深度”则体现在神经网络的层数以及每一层的节点数量。传统的神经网络最多只包含3个层次，结构的简单决定了它能够运行的功能相当有限。再次基础上，深度学习采用由包括输入层、多个隐藏层和输出层组成的多层网络，这种分层结构是深度学习模仿人类大脑的核心结构特征。

要介绍深度学习的原理，就不得不说些题外话。1981年，两位神经生物学家大卫·胡贝尔（David Hubel）和托尔斯滕·魏泽尔（Torsten Wiesel）连同另一位科学家分享了诺贝尔医学奖，他们二位的主要贡献在于“发现了视觉系统的信息处理方式：可视皮层是分级的”。1958年，胡贝尔和魏泽尔在美国的约翰霍普金斯大学开展关于瞳孔区域与大脑皮层神经元的对应关系的研究。他们给小猫展现形状和亮度各不相同的物体，并改变每个物体放置的位置与角度。在这一过程中，小猫的瞳孔感受不同类型和不同强度的刺激，小猫的后脑上则被插入电极，用来测量神经元的活跃程度。

这一实验的目的是验证一个假设：位于后脑皮层的不同视觉神经元与瞳孔感受到的刺激信号之间，存在某种相关性。一旦瞳孔受到某种特定的刺激，后脑皮层的某些特定神经元就会活跃。经过长期枯燥的试验

后，胡贝尔和魏泽尔发现了“方向选择性细胞（Orientation Selective Cell）"：当瞳孔发现了眼前物体的边缘，而且这个边缘指向某个方向时，这种神经元细胞就会活跃。这一发现不仅在生理学上具有里程碑式的意义，更激发了人们对于神经系统的进一步思考，促成了人工智能在四十年后的突破性发展。

方向选择性细胞提示人们：神经-中枢-大脑的工作过程，或许是一个不断迭代、不断抽象的过程。人眼处理来自外界的视觉信息时，遵循的是这样的流程：首先提取出目标物的边缘特性，再从边缘特性中提取出目标物的特征，最后将不同的特征组合成相应的整体，进而准确地区分不同的物体。在这个过程中，高层的特征是低层特征的组合，从低层到高层特征变得越来越抽象，语义和意图的表现就越来越清晰，存在的歧义越来越少，对目标物的识别也就越来越精确。

深度学习在功能上受启于大脑视觉系统中感受视野特征的方式。在深度学习中，这个过程被利用多个隐藏层进行模拟：第一个隐藏层学习到“边缘”的特征，第二个隐藏层学习到的是由“边缘”组成的“形状”的特征，第三个隐藏层学习到的是由“形状”组成的“图案”的特征，最后的隐藏层学习到的是由“图案”组成的“目标”的特征。当然，这样的识别思想不只适用于视觉信息的处理，对其他类型的信息同样适用。

2006年，加拿大多伦多大学教授、机器学习领域的泰斗辛顿在国际权威学术期刊《科学》上刊文，深度学习就此闪亮登场。辛顿的文章表达了两个主要观点：首先，具备多个隐藏层的人工神经网络（也就是深度学习）具有优异的特征学习能力，习得的特征能够实现对数据更加本质性的刻画，有利于对数据的可视化或分类；其次，深度学习在训练上的难度可以通过“逐层初始化（Layer-wise Pre-training）”来有效克服，逐层初始化则可以通过无监督学习实现。

与深度学习相对应的是浅层学习（Shallow Learning），其局限性在于有限样本和计算单元情况下对复杂函数的表示能力有限，针对复杂分

类问题其泛化能力受到一定制约。深度学习克服了浅层学习的弱点，通过深层非线性网络结构实现复杂函数逼近和表征输入数据分布式表示，展现出强大的从少数样本集中学习数据集本质特征的能力。学习特征的过程可以被视为特征空间变换的变换过程，通过逐层特征变换，将样本在原空间的特征表示变换到一个新特征空间。这样的变换能够有效去除不同特征之间的相关性，从而使分类或预测更加容易。

图3-6 乔弗·雷辛顿

深度学习的另一个主要优势是能够从海量数据中进行特征的自动提取。在浅层学习中，依赖先验知识的手工设置特征处于统治地位，这类特征的设计中只允许出现少量的参数，设计出的特征的不变性与可区分性也远非最佳。可深度学习可以从大数据中自动学习特征的表示，其中可以包含成千上万的参数，手工设计出有效的特征是一个相当漫长的过程。回顾计算机视觉发展的历史，往往需要五到十年才能出现一个受到广泛认可的好的特征。而深度学习可以针对新的应用从训练数据中很快学习得到新的有效的特征表示。

深度学习虽然通过特征的自动提取将人从手工特征设计中解放了出来，但目前在神经网络架构中，网络层数、每层神经元的种类和个数、训练算法参数等超参数可能对学习结果有着决定性的影响。这些超参数

的设置和调节，仍然高度依赖人的经验。自动网络结构学习和超参数调节是深度学习从技术走向科学的必由之路。此外，深度学习从原始自然信号中提取特征完成任务的过程是个“黑盒子”，缺乏可解释性，类似于哺乳动物的低级认知功能。与之相对，基于抽象符号和规则的逻辑推理作为人工智能的早期方法，虽然能部分模拟人的高级认知功能，却和现有的神经网络框架“水火不容”。如何把深度学习过程和人类已经积累的大量高度结构化知识融合，发展出逻辑推理甚至自我意识等人类的高级认知功能，是下一代深度学习的核心理论问题。

第四章　得数据者得天下
——智能思维方式的革命

在七卷本的科幻巨著《基地（Foundation）》系列中，首次出现了心理史学（Psychohistory）的概念。心理史学由小说中的数学家哈里·谢顿（Hari Seldon）受物理中电子云的理论启发所提出：在电子云中，微观层面上单个电子的运动是无规律的随机运动，但宏观层面上大量电子的运动却能够精确描述。谢顿将这一理论应用于未来社会，通过将所有个体的行为进行综合来实现大尺度趋势的分析，可以准确地预测出社会的发展走向，甚至进一步在关键点导入适当的变数，进而改变以后社会发展的可能途径。当然，心理史学的实现有两个前提条件：

作为研究对象的人类，总数必须达到足以用统计的方法来加以处理。

研究对象中必须没有人知晓本身已是心理史学的分析样本，即必须保证研究对象的随机性和自发性。

虽然只是科幻作品中的虚构，但“心理史学”的概念在现实中不乏拥趸。2008年诺贝尔经济学奖得主保罗·克鲁格曼（Paul Krugman）就是心理史学的大粉丝。在那个没有学校开设心理史学的年代，克鲁格曼转而钻研经济学，并在自由贸易与全球化进程等领域中成就斐然。

在不具备专业背景的大众眼中，如果说人工智能的概念尚且是阳春白雪，曲高和寡，那么“大数据（Big Data）”这个术语近些年来就可以

称为飞入寻常百姓家的“下里巴人”了。一方面，谷歌、微软、百度、腾讯等国内外诸多互联网企业都将业务重点从技术转向数据，逐渐开始相关的产业布局；另一方面，我国政府对大数据产业的发展也越发重视，上至中央下到地方都通过各种政策鼓励发展大数据产业链。那么，大数据究竟有何魅力，让学术界和产业界都趋之若鹜呢？

工业时代到信息时代
——世界观的重构

从工业时代到信息时代的转变，是从机械思维到数据思维的转变。

所谓机械思维，是指建立在思辨的逻辑推理基础上思维方式。借助机械思维，人类从真实的世界中抽象出无需证明的最基本的公理，再通过因果逻辑由公理推导出各种基本定理，最终在基本定理的基础上构建起富丽堂皇的科学宫殿。机械思维最早的成就是由古希腊数学家欧几里得（Euclid）创立的公理化体系的几何学。事实上，几何学并非欧几里得的首创：在尼罗河流域的古代埃及，幼发拉底河和底格里斯河流域的美索不达米亚以及长江和黄河流域的古代中国，其孕育的文明中就已经包含了几何学的基本知识。但这些文明对几何学的认识还仅限于具体现象的观察，而没能上升到抽象规律的推演——因此这些文明中对几何的认识并没有形成体系，也就不能称之为几何“学”。

欧几里得的贡献正是把散落的珍珠串成了美丽的项链：他首先总结出5条相互独立的不证自明的公理，其中的任何一条都不依赖于其他公理而存在，也无法依据其他公理推导出来。通过直接使用公理和定义作为前提证明，或是间接的通过已经直接证明的定理的证明，欧几里得推导出了所有的几何学定理，由此建立起欧氏几何学的大厦。公理化体系几何学被写在欧几里得的巨著《几何原本（Elements）》之中，为现代数学乃至整个科学的发展奠定了坚实的基础。

图4-1 尼德兰画家尤斯图斯（Justusvan Gent）所作的欧几里得画像

在欧几里得去世约两百年后，古希腊最伟大的天文学家克劳狄乌斯·托勒密（Claudius Ptolemy）将欧几里得的方法论应用到天文学上，建立起一套完整、严格而且相当精确的描述天体运动规律的理论体系——地心说。虽然在我们常规的思维中，地心说总是与教会的强权和愚昧挂钩，支持日心说的科学家们——波兰天文学家尼古拉斯·哥白尼（Nicolaus Copernicus）、意大利数学家乔尔达诺·布鲁诺（Giordano Bruno）和意大利科学家伽利略·伽利雷（Galileo Galilei）——遭受迫害的悲惨境遇甚至使托勒密本人也染上了一丝暴君的色彩。但无辜的托勒密实在是躺着中枪：要知道布鲁诺在罗马鲜花广场因捍卫日心说而遭受火刑时，托勒密已经去世一千五百年有余了。

拨开历史偏见的迷雾，作为科学学说的地心说绝对可谓精品，是机械思维的精妙产物。在建立地心说的过程中，托勒密使用的方法论可以

概括为“通过观察获得数学模型的雏形，再利用数据来细化模型”。在研究天体运行的过程中，托勒密将欧几里得和另一位古希腊数学家毕达哥拉斯（Pythagoras）的数学思想与上百年来观测得到的天文数据相融合，将各种天文现象的共性用最基本的元模型——圆形运动来描述。托勒密仅仅通过圆这种基本形状，以及大小不同的多个圆形的相互嵌套，就把当时人们已知的天体运动的规律描述得一清二楚。至于他提出的为什么是地心说而不是日心说，原因在于这最符合人们看到的现象——日月星辰都是从东边升起，西边落下。即使一千多年之后哥白尼提出了日心说，也仅仅是因为把托勒密坐标系的中心从地球移到太阳，就可以简化天体运动的模型，其思维方法则与托勒密毫无二致。

科学界中机械思维的最后一位集大成者正是英国物理学家艾萨克·牛顿爵士（Sir Issac Newton）。在巨著《自然哲学之数学原理（Philosophiae Naturalis Principia Mathematica）》中，牛顿用力学三定律和万有引力定律这几个简单而优美的公式破解了宇宙中万物运动的规律，还用微积分的概念把数学从静止的变量拓展为运动变化的函数。牛顿通过自己的伟大成就宣告了科学时代的来临，作为思想家，他让人们相信世界万物是运动的，而冥冥之中支配这些运动的规律既是确定的，又是可以被认识的。同时，牛顿还指出正确的规律通常具有简洁的形式，这与东方哲学中大道至简的思想不谋而合。在牛顿之后，英国物理学家詹姆斯·焦耳（James Joule）使用简单的公式描述了能量守恒定律，英国物理学家詹姆斯·麦克斯韦（James Maxwell）又用四个简明的公式概括了看不见摸不着的电磁世界的全部规律。这些都是机械思维在科学中的重要成果。

牛顿的种种成就奠定了他史上最有影响力的人物之一的地位。去世后，牛顿被葬在伦敦威斯敏斯特大教堂，其陵寝规格超过了任意一位英国君主。著名英国诗人亚历山大·波普（Alexan-der Pope）对牛顿一生的贡献给出了精当的评价：

天不生牛顿，万古如长夜。

虽然在工业时代，人类社会所取得的进步大部分得益于机械思维，但是到了信息时代，它的自圆其说遇到了越来越多的困难。一方面，并不是所有的规律都可以用简单的形式体现；另一方面，很多情况下明显的因果关系也并不存在。20世纪初量子力学的诞生与发展迫使人们接受了微观世界这个全新的观察视角，同时也不得不承认不确定性才是世界的本质。自此，机械思维完成了它伟大的历史任务，不确定性观念下的信息论开启了认识世界的全新方式。

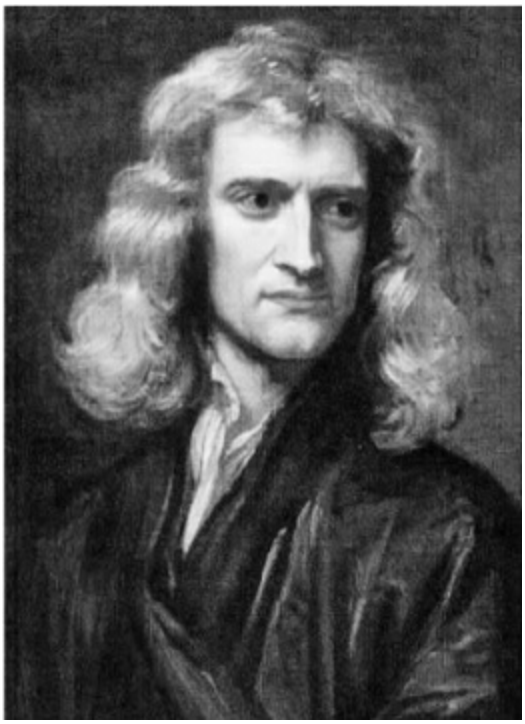

图4-2 英国画家戈德弗雷·内勒爵士（Sir Godfrey Kneller）所作的牛顿画像

随着我们对世界的认识的不断深入，影响事物发展变化的变量也一个个走进视野。如果把影响事件方方面面的因素都纳入考虑之中，这类变量的数目就会多如恒河沙数，已经无法通过简单的办法或者公式算出结果，因此我们宁愿人为地把它们归为不确定的一类，或是做出忽略某些次要因素的必要假设。虽然在实际的火箭发射中必须要考虑到气温水平、湿度水平等诸多影响因素，以确保发射的万无一失，但如果在高考

或者中考中遇到类似的题目，还是可以放心大胆地把空气阻力等因素忽略不计，用简单的牛顿定律来解决问题。

此外，科学研究也在不断证实，不确定性就是客观世界的本质属性。在宏观世界中，行星运动的速度和位置是可以被精确计算的。但是在微观世界里，电子在围绕原子核做高速运动时，它在某个时刻的位置和运动速度不可能同时被准确地测定，自然也就不能描绘出它的运动轨迹了。这样的结果不取决于测量仪器的精度，而是由量子力学中的基础性原理——德国物理学家沃纳·海森堡（Werner Heisenberg）所提出的测不准原理决定的。除此之外，近年来关于无漏洞的贝尔不等式的实验进展也在不断说明，这个世界上不存在任何的隐变量，它本身就是以概率性的方式在运行——换句话说，上帝还真就掷骰子。

不确定性的世界只能使用概率模型来描述，这促成了信息论的诞生。1948年，供职于美国贝尔实验室的物理学家克劳德·香农（Claude Shannon）发表了著名论文《通信的数学理论（A Mathematical Theory of Communication）》，给出了对信息这一定性概念的定量分析方法，标志着信息论作为一门学科的正式诞生。在信息论中，香农以“信息熵”的概念解决了对单个信源的信息量和通信中传递信息的数量与效率等问题做出了解释，并在世界的不确定性和信息的可测量性之间搭建起一座桥梁。

图4-3 克劳德·香农

虽然信息论的首要目的只是建立关于通信的科学理论，但它作为方法论的影响力已经渗透到生活的每个角落之中。与以确定性为基础的机械思维截然相反，信息论完全是建立在不确定性基础上，消除不确定性的唯一方法就是引入信息。这正是信息时代所带来的思维变革：大量机械思维无能为力的问题都可以通过转化为信息处理问题而解决。而大数据的出现，意味着信息时代最有力的工具已经悄然到来，更意味着信息时代的下一次进化。

知其然，而非所以然
——信息到数据的认知变革

虽然作为术语的“大数据”近来才受到人们的高度关注，但在概念上它并不新鲜。著名的《二十四史》实际上就是对我国社会发展的大数据记录。1980年，著名的美国未来学家阿尔文·托夫勒（Alvin Toffler）在其著作《第三次浪潮》中，就已经提及大体量数据对信息技术乃至未来社会发展的影响，但在近四十年前，技术条件的限制使这样的观念显得过于超前。随着宽带通信技术、移动互联网技术和物联网技术的发展，数据正在以前所未有的速度疯狂涌现，这也给大数据的发展提供了物理基础。2008年9月，国际知名学术期刊《自然》推出了名为“大数据”的封面专栏，这意味着主流学术界对大数据的认可与关注。学术界的认可也影响到了工业界与商业界，大数据迅速成为互联网技术行业中的热门词汇。

在作为物理概念的“大数据”的基础上，世界著名的管理咨询公司麦肯锡公司（McKinsey Company）进一步提出了作为商业概念的“大数据”。麦肯锡公司从各类网站上记录的个人海量信息中敏锐地发现了潜在的商业价值，于是投入大量人力物力进行调研，在2011年6月发布了关于大数据的报告“麦肯锡报告”，对大数据的影响、关键技术和应用领域等都进行了详尽的分析。麦肯锡报告得到了金融界的高度重视，使大数据受到了全社会各行各业的关注。

图4-4 大数据登上《自然》封面

2012年，英国牛津大学教授维克托·迈尔-舍恩伯格（Vik-tor Mayer-Schornberger）出版了学术专著《大数据时代》，在书中提出了一系列颇具前瞻性的洞见。舍恩伯格在书中指出，大数据带来的信息风暴正在变革我们的生活、工作和思维，开启重大的时代转型，并为人类的生活创造前所未有的可量化的维度。

说到这里，我们有必要对大数据的内涵加以阐释。实话实说，大数据这个概念还没有多方公认的权威定义，学术机构、商业机构与公共管理机构只是分别从自己关注的角度对大数据进行描述。但在不同的行业视角下，大数据会被解读出不同的内涵与不同的特征，如果将这些局部特征熔于一炉，大数据的全貌就会逐渐浮现：

大数据是指以容量大、类型多、存取速度快、应用价值高为主要特征的数据集合。

与传统意义上的“小数据”相比，大数据最明显也最本质的特征在于

它的体量，也就是大数据的“大”。“大”之所指不仅仅是数据超大的比特数目，更重要的是数据的全面性与完整性。以前，受数据采集技术与数据分析技术的限制，准确分析海量数据几乎是不可能完成的任务，因此只能通过在全体数据中采集出一部分样本，通过精确分析样本的性质来粗略估计数据整体的特征，这也正是统计学的核心任务。但在大数据炙手可热的今天，我们关注的不再是采样出来的数据样本，而是海量数据本身。这就可以正确地考察细节并进行新的分析，而无需考虑采样偏差所导致的错误结论，也不会错过可能被采样过程忽视而淹没在海量数据中的重要细节。毕竟，能从数据中获得的所有规律，都蕴藏在数据本身之中，而用于分析的数据越多，得到的规律就越准确。

葡萄酒的品鉴是专业性极强的领域，从事这项工作的通常是具有数十年品酒经验的专家。品酒师通过观察葡萄酒的色泽与稠度，嗅闻葡萄酒的香气，品尝葡萄酒的味道来判断这个酒大概来自于哪个酒庄，酿造于什么年份。但是这门基于经验的手艺也有它自己的问题：当品酒师品鉴新酒时，由于葡萄酒储存的时间太短，其真正的品质还没有形成，所以品鉴结果难免流于偏颇。另外，知名品酒师爱惜名誉有如孔雀爱惜羽毛，这种怕出错的心态也会影响到对酒类的鉴赏判断，使品酒师倾向于给出随大流的中庸结果。

难道判断葡萄酒水准的话语权只掌握在品酒师手中？美国普林斯顿大学的经济教授理查德·科万特（Richard Covant）偏偏不信这个邪。作为葡萄酒爱好者，他尽可能多地收集关于葡萄酒产地信息与气候信息的数据，根据这些数据和相应的葡萄酒的质量，科万特得出结论：葡萄酒的品质跟土壤的成分、生长期的平均气温、冬天的降雨量、和收获季节的降雨量等因素有关。根据自己的秘诀，1989年葡萄酒刚一上市，科万特就预测这一年的葡萄酒是世纪佳酿。可仅仅一年之后，科万特又宣称1990年的酒甚至比1989年的还要好！

连续两年号称世纪佳酿，这对任何品酒师来说都是砸牌子的说法，

可科万特就是这么大胆！作为一个外行，科万特对酒的判断不是基于葡萄酒本身，而是生产过程中影响葡萄酒品质的众多天时地利的因素。他可能对葡萄酒的术语一无所知，却能够根据数据作出判断。在习惯的认知方式中，追求的主要目标是线性的、双边的直接因果关系。但是万物之间的联系恐怕比想象的要复杂千万倍，这种联系以多元且非线性方式存在。大数据的出现颠覆了原有的认知模式：认识事物的方式变成了先寻找相关关系，再寻找因果关系。

认知模式的转换反过来也成为审视大数据的慧眼。如何在纷繁复杂的海量数据中提炼出有用的结论呢？方法很简答：从传统的因果分析转向相关性分析转换。相较于统计学中的知其所以然，在大数据时代，只要知其然就已经足够了。当大数据占据我们这个信息社会的中心舞台，传统知识观中的因果律遭到极大的挑战，而相关性则让我们从对过去的理解中解放出对未来的预测，这从本质上改变了数据的利用模式。

从因果性到相关性一个经典的例子就是谷歌对流感爆发的预测。2009年2月，谷歌的研究人员在《自然》发表了一篇论文，预测季节性流感的暴发，在医疗保健界引起了轰动。谷歌对2003年和2008年间的5000万最常搜索的词条进行大数据“训练”，试图发现某些搜索词条的地理位置是否与美国流感疾病预防和控制中心的数据相关。疾病预防控制中心能够跟踪全国各地的医院和诊所病人，但它发布的信息往往会滞后一两个星期，但谷歌的大数据却是发现实时的趋势。

数据往往都是不完美的，拼写错误和不完整短语很普遍。为什么谷歌可以实现这么精准的预测？如果从因果关系看，是因为人感到不舒服，或听到别人打喷嚏，或者阅读了相关的新闻后感到焦虑吗？谷歌不是从这种因果关系去考虑，而是从相关性的角度，去预测一个持续发展的大方向，因为大众的搜索词条处于不断变化之中，外界的一个蝴蝶翅膀的扇动，就会使搜索发生系统的、混沌的变化。谷歌并没有直接推断哪些查询词条是最好的指标。相反，为了测试这些检索词条，谷歌总共

处理了4.5亿个不同的数字模型，将得出的预测与2007年和2008年疾病预防控制中心记录的实际流感病例进行对比后，谷歌公司发现，它们的大数据处理结果发现了45条检索词条的组合，一旦将它们用于一个数学模型，它们的预测与官方数据的相关性高达97%。

关于使用大数据中的相关性提取有用结论的最近一个例子发生在美国国家橄榄球联盟的赛场上。2016年11月8日，一场如火如荼的橄榄球比赛已经进行到第三节，5:21落后的亚特兰大猎鹰队正推进到本方46码线。此时此刻，大数据公司Splunk做出了一个预测：猎鹰队下一步将祭出“霰弹枪阵式”，随后四分卫马特·瑞安将送出一记左侧的短传。随后赛场的形势发展与Splunk的预测如出一辙：猎鹰队果真使用了“霰弹枪阵式”，只不过在最后一传上出现了失误。

Splunk做出这个预测的依据并非依赖于专业的橄榄球从业人员，恰恰相反，这些从事数据分析的极客们可能连橄榄球的规则都不懂。但他们把至少一整年的比赛数据输入计算机，利用计算机来分析不同赛场形势和不同攻守策略之间的联系，从而得出精确的预测。这背后的因果性自然是橄榄球专业人员的技战术设计，但利用相关性也可以得到同样的决断。

海纳百川，有容乃“大”——被量化的世界

英国物理学家开尔文勋爵（William Thomson,1st Baron Kel-vin）曾说过：“当你能够量化你谈论的事物，并且能用数字描述它时，你对它就确实有了深入了解。但如果你不能用数字描述，那么你的头脑根本就没有跃升到科学思考的状态。”

这样的论断在百年后的大数据时代将被奉为圭臬。在海量数据中，量化的价值并不体现狭义的精确定量关系中，而是确定事物背后的运转规律，其出发点不是消除不确定性而是减少不确定性。尤其在大数据时代，分析数据更加追求关联性而非结构性，量化数据也不是非要用数字化去表达，这样的观念变革或许对于数据分析和量化而言是突破性的，而突破点就在于目的性的把握上。正因如此，数据可视化已经逐渐演进为一门独立的学科，它研究的正是如何将数据背后的定量关系直观地展示出来。

在第84届奥斯卡奖评选中，由好莱坞著名编剧阿伦·索尔金（Aaron Sorkin）编剧，金球奖得主布拉德·皮特（Brad Pitt）主演的影片《点球成金（Moneyball）》狂揽六项提名（只可惜全部陪跑……）。这部体育题材影片改编自真实的故事：比利是美国职业棒球大联盟中奥克兰运动家队（Oakland Athletics）的经理。作为一支小本经营的球队，奥克兰运动家无法像财大气粗的豪门纽约扬基一样挥舞钞票开展金元攻势，大肆招兵买马，面对主力纷纷跳槽的窘境，未来的赛季似乎前途渺茫。可一次偶然的机会，比利认识了耶鲁大学经济学硕士彼得，两人对于球队运营的理念不谋而合。比利立即聘请彼得作为顾问，用数学建模的方式，逐渐开始挖掘上垒率的潜在明星，并通过软磨硬泡将他们招致麾

下，并最终上演了人民群众喜闻乐见的簪丝逆袭戏码。

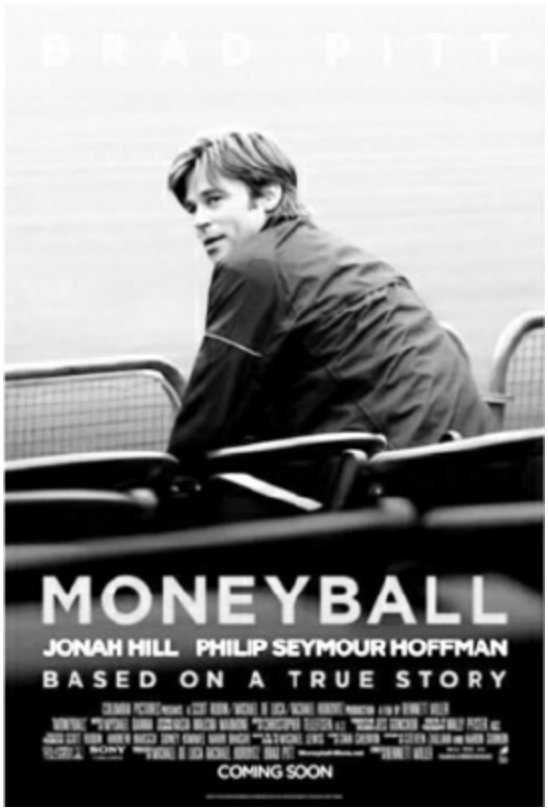

图4-5 电影《点球成金》海报

电影本身的内涵非常丰富，但从数据科学的角度来看，比利所做的事情就是一改老派的教练员基于直觉和经验的球员评价体系，而是对它进行了全方位的量化。棒球本身即是一项强调数据的运动，衡量球员的指标包括打击率、长打率、防御率、胜投数、全垒打数、打点数等数十项指标。可长久以来，棒球界却没能将这些意义非凡的数据转化为球队的战斗力，可谓守着金矿要饭吃。比利正是老旧传统的改造者。现实中，他和同伴建立了号称“棒球统计学（Sabermetrics）”的全新方法，通过统计学的方法将球员能力最大程度地量化，并以量化结果作为衡量球员能力的唯一标准，而非某些基于主观经验的判断。与此配套的是全新的评价体系：让棒球比赛结束的因素是27个出局数，那么“上垒率”就是不二法门，其他诸如“击球率”“盗垒”等华而不实的指标统统都要靠边站。通过这样的方式，比利颠覆了看重球员速度、力量和打击率的传统

思维，挖掘出了决定比赛走势的深层次量化结果，给球队带来了实质性的收益。

大数据基础上的量化与其说是方法的进化，不如说是观念的改变。不经处理的数据本身谈不上价值，而量化才是数据价值提取的核心步骤。只要选择了合适的标准和参考系，万事万物皆可量化。量化是数据价值提取的基础，它能够使很多难以确定的情况变得能够估计和判断，相关的决策与结论才会具备说服力与可操作性。

1965年的诺贝尔文学奖被前苏联作家米哈伊尔·肖洛霍夫（Mikhail Aleksandrovich Sholokhov）以描述哥萨克生活的史诗巨著《静静的顿河》摘得。这部作品以细腻的笔触刻画了哥萨克这一特殊群体在历史漩涡中的生活与命运，是俄罗斯文坛上的璀璨明珠。但在当年美苏争霸的国际形势下，出于各种各样的原因，以苏联著名异见人士亚历山大·索尔仁尼琴（Aleksandr Isayevich Solzhenitsyn）为首的诸多知名人士质疑《静静的顿河》并非出自肖洛霍夫本人之手，而是抄袭了俄国内战中一位白军军官克留科夫的笔记。这一观点随着诺贝尔奖的颁发愈发甚嚣尘上，变成了文坛的一桩悬案。

图4-6 米哈伊尔·肖洛霍夫

在肖洛霍夫获奖20年后，这桩沸沸扬扬的笔墨官司终于尘埃落定。1984年，挪威奥斯陆大学的数学家与斯拉夫研究专家盖尔·克耶萨（Geir Kjetsaa）运用数理统计的分析方法对《静静的顿河》进行了研究，证实了肖洛霍夫是本书的作者。这一成果被克耶萨及其合作者写书出版，轰动一时，克耶萨教授与他的合作者使用乌普沙拉大学的一台IBM370/155电子计算机，对《静静的顿河》与“被抄袭者”克留科夫的一些作品进行比较。比较的方法是对肖洛霍夫和克留科夫的文本分别进行抽样，再编写程序测定句子长度和词汇分布等参数，据此生成对两人写作风格的比较。为了执行对比，所有的原始材料被分为三组：肖洛霍夫的无可争议的作品为第一组，《静静的顿河》为第二组，克留科夫的作品为第三组，研究者则分别研究三组文本的三个重要参数：

第一个参数是作品中出现的不同的词汇数量与总词汇量的百分比统计：三组结果分别65.5%，64.6%和58.9%。显然前两个数据非常接近，并明显高于第三个数据。这表明肖洛霍夫的语言风格更加多变，而

克留科夫偏爱使用重复的词汇。

第二个参数是词汇分布频率：研究者们选取了20个常见的俄文词汇，统计其在作品中出现频率。三组结果分别为22.8%，23.3%和26.2%，体现出与第一个参数同样的趋势。看起来这些词更受克留科夫的青睐。

第三个参数是作品中出现过一次的词汇所占的百分比：三组结果分别为80.9%，81.9%和76.9%。这表明肖洛霍夫的词汇量要高于克留科夫。

在不同文本的比较中，三组参数表现出了一致的趋势，即克留科夫的作品与《静静的顿河》之间存在着显著的统计差异，这部杰作的真正作者更像是肖洛霍夫。这一结论在1999年被证实：《静静的顿河》手稿被发现，其中605页为肖洛霍夫亲笔，另285页他的妻子和姐妹誊写。这也给这段公案画上了一个句号。

虽然克耶萨教授的研究已过去二十年有余，但他解决问题的思路正是大数据量化的思维方式：写作风格本来是虚无缥缈的东西，却可以通过作为载体的文本却是看得见摸得着的，其中体现出来的作者遣词造句的方式也难以伪造。对词语和句法的数理统计无疑就是对写作风格的量化。对四大名著之一的《红楼梦》后四十回的真伪判定也使用了类似的方法。当然，受当年的技术条件限制，克耶萨教授分析的对象只限于抽取出来的文字样本，这将不可避免地给分析结果带来偏差。在大数据处理技术日臻成熟的今天，如果对全部文本进行统计的话，也许会得到更具说服力的结果。

在这个大数据的时代，数据正在从最不可能的地方涌现出来。量化一切是数据化的核心：一串串字符是对文字的量化；数字音频是对声音的量化；各种格式的数字图片是对图形的量化。量化正在不断推进数据化的进程：地图类应用是对地理场景的数据化；形形色色的电商平台上

琳琅满目的商品是对现实物品的数据化；服务网站上各种各样的供需信息是对服务的数据化；微博和论坛是对思想观点的数据化；转发和点赞是对传播的数据化；社交网络是对人际关系的数据化。人和物的一切状态和行为都能数据化，而数据化意味着事务在数据空间里的极易操作，往往由此生发出伟大的创意。

有数据，才有一切
——人工智能驱动力

人工智能离不开深度学习。通过大量数据的积累探索，机器必将在任何单一的领域超越人类。而人工智能要实现这一跨越式的发展，把人从更多的劳力劳动中彻底解放出来，除了计算能力和深度学习算法的演进，大数据更是助推深度学习的高能燃料。离开了大数据，深度学习就成了无源之水、无本之木。

深度学习的实质，是通过构建具有很多隐层的机器学习模型和海量的训练数据，来学习更有用的特征，从而最终提升分类或预测的准确性。从本质上来，深度学习只是手段，特征学习才是目的。为了更加精确地学习特征，深度学习引入了更多的隐藏层和大量的隐层节点；明确突出了特征学习的重要性，也就是说，通过逐层特征变换，将样本在原空间的特征表示变换到一个新特征空间，从而使分类或预测更加容易。与人工规则构造特征的方法相比，利用大数据来学习特征，更能够刻画数据的丰富内在信息。

从实际应用的角度来说，深度神经网络只是一个可以运作的简单大脑，单靠这个简单的大脑还不足以完成深度学习的任务。在医学上有种现象：聋哑儿童由于先天或后天的原因在年幼时丧失了听力，但他们的发声功能通常完好无损，这意味着具备说话的生理条件。可长大后，大部分的聋儿都不会说话，只能发出类似语言的简单音节组合。完好的生理条件并没有进化成语言能力，这是为什么呢？

其原因正是在于语言的能力没有被训练出来。读者不妨回忆自己学习学话的过程：一没有理论学习，二没有题海战术，靠的就是简单的咿

呀学语。幼儿在最初听到任何语言的时候都会蒙圈，不知道说的到底是什么东西，但他们会通过观察出现这些语音信号时的场景图像，来猜测这些词句大概代表的含义，并将图像和语音建立联系。经过多次的重复刺激后，幼儿就会逐渐形成了对这一语音符号的“条件反射”，在大脑语言区的位置形成了脑神经的一个网络结构逐渐构造该语言的语言区，最终实现了用这种语言的语音符号思维的能力。而对于聋儿来说，听觉的丧失使他们无法建立图像和语音之间的联系，也就没有办法形成习得语言所必备的条件反射了。

根据连接主义学派的观点，机器的深度学习借鉴的正是人类的学习，训练的过程也是智能形成的必由之路。如今，大数据就扮演着这一重要的“训练”角色。大数据的飞速发展，让深度学习拥有了无比丰富的数据资源来完成特定功能的“训练”。除此之外，拜发达的传播渠道所赐，大数据还能够产生涟漪效应：千千万万的深度学习用户把与之相关的使用习惯传入已有的数据集合中，新增的数据反过来又能够促进学习的深入。这样的涟漪效应使深度学习不断地进行自身的优化去达到更优的结果。前文中提及的阿尔法狗便是大数据训练出来的硕果：古今中外的海量对局愣是把不懂围棋为何物的算法训练成了独孤求败的高手。

大数据的出现为深度学习的发展提供了前所未有的契机，却也对它提出了更高的要求。工业界一直奉行“大道至简”的原则：在大数据条件下进行机器学习，简单模型会比复杂模型更加有效。可近年来深度学习的惊人进展，促使我们不得不重新思考这个观点。在大数据情况下，也许只有比较复杂的模型，或者说表达能力强的模型，才能最大程度地发掘出海量数据中蕴藏的丰富信息。大数据运用到浅度学习上，只会产生消化不良的后果，只有更强大的深度模型才能从大数据中发掘出更多有价值的信息和知识。

语音识别是一个典型的基于大数据的机器学习问题：其声学建模的训练样本可以达到十亿甚至是千亿级别。要处理这样体量的数据，普通

的神经网络是无能为力的，需要更加复杂的深度神经网络。可在谷歌公司的一个语音识别实验中，研究者发现即使使用深度神经网络进行训练，训练出的模型对训练样本和测试样本的预测也相差无几，这意味着所有的训练都打了水漂，连个响动都没听见。这种违背常理的现象只有一种解释，就是由于大数据里含有的信息维度太过丰富，即使是如深度神经网络一般的高容量复杂模型也处于欠拟合的状态，更不必说传统的高斯混合声学模型了。深度学习模型就像是高效的冶炼机器，没有它就没有办法从大数据这座金矿里提取出金子。

要使机器大脑达到人脑的水准，第一个重要的步骤就是获取信息。信息既可以通过搜索引擎直接抓取，也可以通过记录用户的搜索历史获得。当然，孤立的信息是没有任何用处的，机器大脑还要挖掘其中的各种关联，作为行动的指导。这个过程很难由机器主动完成，现阶段唯一的途径是通过搜索引擎的用户的反馈实现：当用户搜索某个关键词后对某个网站点击增加，就会自动增加这个关键词与该网站的关联，不断地寻找最优算法，让用户直达最优结果。

事实上，不只是语音识别或是图像识别这类专门的应用，真正的人工智能也应当基于大数据而诞生，并基于大数据不断进化。通过对海量的搜索和其他相关操作进行关联性的提取与分析后，机器大脑就能够找出在发生某个特定事件时，绝大多数人类的行为模式，并以这种模式和人类进行交互，使人以为对面真的是一个人。在现有的技术条件下，这可能是人工智能的终极形态：一个没有鲜明个性的“人”，一个群体意志的产物。

第五章　我，机器人
——人工智能的终极载体

机器人学三定律

第一定律：机器人不得伤害人，也不得见人受到伤害而袖手旁观。

第二定律：机器人应服从人的一切命令，但不得违反第一定律。

第三定律：机器人应保护自身的安全，但不得违反第一定律和第二定律。

引自《机器人学指南》第56版

图5-1 艾萨克·阿西莫夫

从《终结者》中冷酷无情的T-800到《人工智能》中多愁善感的小朋友大卫，从《变形金刚》里无所不能的威震天到《她》中温柔体贴的萨曼莎，机器人已经成为对人工智能物理实体的终极想象。可人类对机器人的最初期待只是让它们成为生活助手，把我们从重复单调的劳动中解放出来。著名科幻小说家艾萨克·阿西莫夫（Issac Asimov）在科幻小说《我，机器人》中提出了著名的机器人学三定律，但如果机器人无所不能，世界上还会有属于人类的位置吗？在人工智能和机器人技术之间，未来到底是乌托邦还是索多玛？

思考能力的进化
——语音助手与无人驾驶

今天，机器人与人工智能这两个概念已经是你中有我，我中有你。可出人意料的是，机器人的英文单词robot却是在一个世纪之前由一位文学家创造的。1920年，捷克斯洛伐克作家卡雷尔·恰佩克（Karel Capek）在他的剧本《罗萨姆的机器人万能公司》中，根据捷克语单词robota（意为“劳役、苦工”）和波兰语单词robotnik（原意为“工人”），创造出“机器人”这个单词。由此可见，人类对机器人的定位只是一种基于机械设备的劳动工具，跟智能原本是八杆子打不着的。

图5-2 《罗萨姆的机器人万能公司》剧照：反抗的机器人

虽然机器人一词的出现和世界上第一台工业机器人的问世都是二十

世纪的事情，但人类对机器人的追求却源远流长。根据西周时期的记述，当时的能工巧匠偃师给周穆王献上用动物皮、木头、树脂制造出的美女伶人——无疑是今天各式各样类人机器人的老祖宗。更牛的是，这位美女不仅能歌善舞，甚至还有六欲七情。这样的木头伶人固然是寓言中的幻想，但它从一个侧面反映出当时的科技发展水平，也是中国最早记载的木头机器人雏形。

中国古代关于机器人的记述可不光这一种。著名的科学典籍《墨经》中曾记载，木匠祖师爷鲁班曾经使用竹子和木头作为原料制造出一只木鸟，在空中连续飞行三天三夜；而根据《三国演义》中的说法，鞠躬尽瘁死而后已的诸葛亮不光韬略出众，工程技术水平也相当了得：他设计的具备传动装置的木牛流马可称为最早的陆地军用机器人，运输粮食时本来大步流星，但舌头一扭便纹丝不动，只留下劫粮成功满心欢喜的魏军大眼瞪小眼。

机器人的设计并非中国人的专利。看过《达芬奇密码》的人必然对列奥纳多·达芬奇（Leonardo da Vinci）的惊世之才赞叹不已，历史上的达芬奇其人也的确是个百科全书式的人物，对于机械设计尤其在行。15世纪，达芬奇在人体解剖学的知识基础上利用木头、皮革和金属外壳设计出了机械版的装甲兵。根据记载，这个机器人以齿轮作为驱动装置，肌体间连接传动杆，不仅可以完成一些简单动作，内部的自动鼓装置还能以阵阵鼓声提振士气。三百年后，瑞士的钟表匠人使用凸轮控制和弹簧驱动的原理设计出了三个真人大小的机器人——写字偶人、绘图偶人和风琴偶人，这些多才多艺的偶人直到今天还保存在瑞士纳沙泰尔市的博物馆中。

在20世纪以前，机器人的发展动力还来源于机械控制，因而只能说是技术，而称不上是科学。这一局面在1948年被打破：天才学者维纳（人工智能行为主义学派的奠基人，参见第一章）建立起新型学科“控制论（Cybernetics）”并出版了同名巨著，阐述了机器中的通信和控制

机能与人的神经、感觉机能的共同规律。在此基础上，维纳深入探讨了机器与人的统一性，揭示了机器通过反馈控制模拟人类的可能性，这不仅成为人工智能中行为主义学派的理论基础，也为机器人的发展开拓出一条新路。

自此，计算机技术走进了机器人的身体，使机器人的发展步入了新纪元，并催生了机器人产业：1954年，工业机器人先驱乔治·德沃尔（George Devol）创造了世界第一台可编程机器人“尤尼梅特（Unimate）"，并借此东风与他的商业伙伴约瑟夫·恩格尔伯格（Joseph Engelberger）创立了全世界第一家机器人公司Unimation，打响了机器人产业的第一炮。尤尼梅特的核心技术是借助伺服技术控制机器人的关节，利用人手对机器人进行动作示教，机器人能实现动作的记录和再现。它采用了分离式固体数控元件，并装有存储信息的磁鼓，具有更好的通用性和灵活性。与此同时，美国机床制造公司也研制出了工业机器人“万能搬运（VERSAtile TRANsfer）"。万能搬运主要用于机器之间的物料运输、采用液压驱动。该机器人的手臂可以绕底座回转，沿垂直方向升降，也可以沿半径方向伸缩。尤尼梅特和万能搬运是世界上最早的商业化工业机器人，代表了当年可编程机器人的最高水平。

时间进入20世纪60年代后，传感器技术被应用在机器人产业中，提升了机器人的可控制性。带有触觉传感器的数控机械手被引入机器人中，这种机械手能够自主识别块状材料并将其堆叠起来，而无需人工干预。约翰霍普金斯大学则研制出机器人“怪兽（Beast）"。怪兽已经能通过声纳系统、光电管等装置根据环境校正自己的位置，这使它在合理的非结构性环境中具备了自适应的特性。斯坦福大学则更进一步，开发出有手（机械手）、眼（摄像头）和耳（拾音器）的机器人“摇摆（Shakey）”。摇摆能够“看见”散放在桌面上的方块，识别语音指令并按指令进行操作。在这一时期，几个工业化国家竞相开展具有视觉触觉和优异的操控性能，能避障、钻洞、爬墙甚至水下移动的各种智能机器人的研究工作，并开始在海洋开发、空间探索和核工业中试用。在20世

纪80年代中期，机器人的技术水平越来越完善，在各类工业生产中广泛普及，机器人应用制造业已成为发展最快和最好的经济部门之一。

可编程能力与传感技术使机器人具有感觉、识别、推理和判断的能力，能够适应外界条件变化并自适应地对自身工作进行调整，达到了智能的程度。但归根到底，这类智能机器人运行的规则仍然由人类规定，在特定的环境下如何调整也要遵循预先设定的原则。由此，机器人下一步的发展目标就是具备更高级别的智能：机器人自己通过学习，总结经验来获得修改程序的原则。这种机器人已拥有一定的自动规划能力，能够自动安排自己的任务，可以在无人值守的情况下完全独立的工作。伴随着这一潮流诞生的则是智能机器人的全新形态。

2011年10月，苹果公司在其手机操作系统iOS中首次集成了语音助手，Siri就此闪亮登场。由于Siri的早期版本功能较弱，调戏Siri也就成了苹果手机用户喜闻乐见的谈资，但语音助手作为机器人的全新模式，其的流行就此势不可挡。2012年7月，谷歌公司推出了自己的语音助手Google Now.2014年2月，不甘落后的微软公司也推出了自己的语音助手小娜（Cortana），并嵌入安装Windows操作系统的计算机和手机中。借助微软自身深厚的技术功底，Cortana实现了对语音的较高识别率和与系统功能的深度集成，给微软用户带来了不少便利。很快，微软又趁热打铁推出了跨平台的聊天程序小冰，这个“不良少女”所引起的槽点满满的轩然大波算是语音助手发展历史中一个有趣的注脚。

Siri也好，小娜也罢，其本质都只是一团代码，与传统意义上的机器人搭不上一点儿关系。可他们又确确实实是机器人——不仅形态全新，而且功能全新。语音助手的简化形式——俗话说的聊天机器人，也就是前面所说的小冰——具有对话和在对话中学习两种基本功能。对话的原理是先从对方提问中提取关键词，再到网络中检索匹配度高的答案，以此回答对方；而从对话中学习，是在过往对话中提取有效对话增加到自己的数据库中。事实上，这类聊天机器人更像是个搜索工具：把

人类的对话输入作为关键词在网络上搜索对应的回复。目前，语音助手在事务性指令上处理的效果较优，但聊天机器人依然是答非所问和不知所云的代名词。可这样的结果恰恰反证了方式的先进性：语音助手并非按照预设的规则回答问题，而是完全以自主学习的方式实现对自然语言的理解，这与传统的工业机器人有本质区别。虽然受技术条件限制，现有程序还不能在大数据的研究中习得深层次的判断能力和理解能力，但在原理上并不存在不可逾越的障碍。而对于传统的工业机器人来说，自主学习显然是不可完成的任务。

关于机器人思考能力的进化更直观更形象的实例，是无人驾驶汽车的出现。

早在上世纪的八九十年代，德国慕尼黑和意大利帕尔玛就出现过无人驾驶的演示。无人驾驶在诸如换档、变速等技术上的困难已经被完全克服，换句话说，在无人的环境下无人驾驶一点儿问题都没有。可是在全天候全路况的实际条件下，早年间的无人驾驶技术就会变身马路杀手。这也说明了为什么无人驾驶由互联网巨头谷歌，而非奔驰、奥迪等传统车企取得突破。由于本书并非技术类书籍，因而无人驾驶的技术细节在本书中不做赘述，而是用一些数据代替：截至2015年6月，谷歌开发的无人驾驶汽车已行驶超过160万公里，相当于成年人大约一辈子的驾驶公里数；截至2016年2月，无人车经历了14次交通事故，其中13次是由人类司机的行为导致。只有一次是障碍避让导致了撞车。

图5-3 由凌志RX450h改装的谷歌无人驾驶汽车

从机器人的发展历程来看，谷歌无人驾驶汽车具有里程碑式的意义。人类的自然科学从一开始就建立在严密的逻辑和规则之上，机器人也是从空间上时间上都可以精确执行的任务开始。在这个层次上，人类已经把各种内部外部的影响因素做好了建模，机器人只是求解这一模型的数学工具。随着实现的功能的复杂化，机器人的结构也变得越来越精细，但它不具备、也不需要具备基本的认知能力。谷歌汽车的出现实质上就是对建立机器人认知能力的尝试。复杂的人类世界是不可能用确定的模型描述的，要适应这样的环境，机器人就必须自适应地理解环境，首先是从大量实例中学习必要的规则，再根据学习的规则形成判断的思维，最后利用这思维指导实际的决策。与传统的工业机器人相比，谷歌汽车在工程上的突破也许有限，但它却有希望成为第一个在复杂环境中替代人类劳动的机器人，这无疑意义深远。

无论是语音助手还是无人驾驶，都有很长的路要走，可它们已经指明了机器思考能力进化的道路。

乌合之众还是有血有肉
——集群智能

“在第十一届中国国际航空航天博览会上，我国第一个固定翼无人机集群飞行试验以67架飞机的数量打破了之前由美国海军保持的50架固定翼无人机集群飞机数量的纪录。‘集群智能’作为一种颠覆性技术，一直被军事强国视作军用人工智能的核心，是未来无人化作战的突破口。把无人机群作为一个整体来控制，对未来无人机作战及应用方面有广阔的前景。与单机作战平台相比，无人机群在作战时具备功能分布化、体系生存率高、作战成本低等优势。在对抗过程中，当部分个体失去作战能力，整个集群仍可以继续执行作战任务。”

上面的文字节选自新华社的一篇报道，它展现出我国国防科工的重大进展，也涉及到“集群智能”这一人工智能中的重要概念。

在巴西的亚马逊雨林中，几十万只行军蚁——已知的行为最简单的生物——正在行进。用现在时髦的话说，这是一支去中心化、自组织的大军。在这个蚁蚁平等的团体中，单个蚂蚁几乎没有视力，也不具备什么智能，可聚集成团体的它们组成了扇形的团状，一路风卷残云地吃掉所有能吃掉的，带走所有能带走的。高效的它们只需一天就能摧毁雨林里一个足球场面积内的所有食物。到了夜间，蚂蚁会自发形成一个球体，保护起蚁后和幼蚁，天亮后又各就各位，继续行军。

图5-4 2017年美式橄榄球超级碗比赛中，由无人机集群组成的美国国旗图案

快速飞行的蝙蝠群在狭窄的洞穴中互不碰撞，大雁群在飞行时自发地排列成人字形、海洋鱼群通过几何构型充分利用水流的能量……这些自然界中的集群行为早早就吸引了人类的注意。在由大量数目的生物个体构成的群体中，不同个体之间的局部行为并非互不相关，而是互相作用互相影响，进而作为一个整体性的协调有序的行为产生对外界环境的响应。生物群体正是通过个体行为之间的互动获得更积极的响应方式，达到“整体大于部分之和”的有利效果——一百只行军蚁在一起只会横冲直撞，一百万只行军蚁在一起却整齐划一。

实现集群智能的智能主体必须能够在环境中表现出自主性、反应性、学习性和自适应性等智能特性，但这并不意味着群体中的个体都很复杂。集群智能的核心是由众多简单个体组成的群体能够通过相互之间的简单合作来实现某一功能，完成某一任务。其中，“简单个体”是指单个个体只具有简单的能力或智能，而“简单合作”是指个体与其邻近的个体进行某种简单的直接通信或通过改变环境间接与其他个体通信，从而可以相互影响、协同动作。

与高大上的深度学习不同，集群智能不需要汪洋浩瀚的物理数据，也不需要艰深晦涩的数学算法——蚂蚁和大雁会计算微积分吗？集群智能的基础只是作用于个体的运行准则和作用于整体的通用目标——通常

还都很简单。可正是数量足够庞大的简单规则在孕育出了整体意义上的高级智能，这之中除了对量变引发质变的哲学观点的验证，是否还隐藏着“一生二，二生三，三生万物”的神圣感呢？

事实上，在人类的群体中也存在着类似的集体行为，只不过不是以“智能”的方式为人所知，而是臭名昭著的“盲从”——一群行人里，只要有一个人按照指示向着某个目标前进，就算其他人并不知道目标是什么，甚至不知道有目标的存在，这一群人最后也还是会逐步被带往目标方向。作为一种社会心理学现象，集体行为在法国社会学学者古斯塔夫·勒庞（Gustave Le Bon）的经典著作《乌合之众》和美国社会学家埃里克·霍弗（Eric Hoffer）的代表作《狂热分子》中都有精彩的解释与分析。“在人类社会中，个人一旦形成群体，便智商尽失……甚至乐于牺牲自己的个人利益以达到群体目标。”引用于《乌合之众》的这句话可称为真理，因为把它从社会学领域套用到生物学领域，根本不会产生一丝违和感。

图5-5 古斯塔夫·勒庞

从抽象的角度来说，群体行为是大量自驱动个体的集体运动，每个自驱动个体都遵守一定的行为准则，当它们按照这些准则相互作用时就会表现出上述的复杂行为。群体本身不具备中心化的结构，而是通过个体之间的识别与协同达成稳定的分布式结构。这个分布式结构随着环境的变化，以自身为参考系不断趋于新的稳定。集群智能（Swarm Intelligence）正是群居性生物通过协作表现出的自组织与分布式的宏观智能行为，具有如下的特点：

分布式的特点决定了集群智能具有较强的环境适应能力和较强的鲁棒性，不会由于若干个体出现故障而影响群体对整个问题的求解。

群体中的个体间通过影响环境实现间接通信，这决定了集群智能会随着个体数目的增加而增强，即具有较好的可扩展性。

群体中的个体所遵循的行为规则非常简单，这决定了集群智能实现简单。

群体表现出来的智能是在简单个体的交互过程涌现出来，这决定了集群智能具有自组织性。

由群聚而产生的集群智能，能否应用在机器人上呢？从特征上来说，简单的机器人和诸如蚂蚁之类的低智能动物并无二致，只要设计出合适的互动算法，大自然创造的集群智能没有理由不能出现在机器人上。截至目前，规模最大的机器人集群是由哈佛大学的研究人员开发的由1024个机器人组成的Kilobot.这个大规模的机器人群体能够在没有外界干预的情况下，自组织地形成复杂的二维形状。运行在这些小机器人上的是一种包含三种初等集体行为的自组织算法。这三种行为包括：机器人沿着群体的边缘移动，机器人能够产生梯度信息并发送给其他机器人，机器人通过通信和距离检测来定位。群体中的每个机器人都具有相同的算法程序，并且知道期望的二维形状。在机器人实体设计和算法设计的基础上，研究人员用大规模机器人群体实现了自然界中生物所具有的通过集体行为形成复杂结构的能力。

如果说Kilobot只是用于科研的原型设计，那么无人机集群的智能化可是实打实的落地应用，正如节首的新闻报道所指。无人机集群采取一系列由大量分散的低成本系统协同工作，这与投资开发造价昂贵、技术复杂的多任务系统策略无疑有云泥之别。针对不同类型的工作目标，无人系统集群可利用混合搭配的异构优势，通过集群自行分解、分段作业无缝完成任务低成本、高效率的完成工作。在通信与决策上，分布式的无人机集群利用自组网技术可以实现信息的低能耗高速共享，形成具备自愈合功能的、执行信息搜集和通信中继等行动的主动响应网络。无人机集群数据链网络能够支持冗余备份机制和具有一定的自愈能力以提供可用性保证，集群网络能够监控已建立连接，具备应对意外中断的自动恢复能力，同时具备拥塞处理和冲突应对能力。

集群控制算法是无人机集群的关键要素：要实现相互间的协同就必须确定无人机之间逻辑上的信息关系和物理上的控制关系。集群中的体系结构设定可以将无人机集群和人工控制有机结合，保证无人机集群中信息流和控制流的畅通，为无人机之间的高效协同提供框架。除此之外，集群控制算法还有具备良好的可扩展性，即集群中个体数目的变化不会影响控制系统的整体结构。

无人机集群中的另一个关键问题是通信网络的设计。在蜂群和蚁群等动物集群中，通信是通过一种名为“信息素”的物质实现的。在分布式、自组织的无人机集群中，每一个个体都是一个网络结点，其空间的分布也就决定了网络的的拓扑，而网络拓扑自然会影响通信性能。在一定的通信拓扑及性能下，让无人机集群根据所执行的任务智能分配通信资源，提高通信质量，是集群技术的难题之一。要解决这个问题通常需要人为的网络控制策略作为导引。此外，无人机在实际飞行中如果遇到突发状况，必须进行航迹重新规划，以规避威胁，避免“飞蛾扑火”的尴尬。为满足协同工作时的时效性，重新规划所采用的算法必须满足实时、高效的要求。常见的解决方案是根据蜂群算法领域搜索的特点，以参考航迹的突发威胁作为领蜂航迹，跟随蜂仅在参考航迹的突发威胁段进行领域搜索，而不需要对整条航迹进行搜索，由此可以快速获得修正航迹段，并替换原突发威胁航迹段，整个飞行过程中，无人机根据获得的威胁信息，不断修正参考航迹，直至达到目标节点。

在军事领域中，无人机集群有着广阔的用武之地。在第一次世界大战期间，英国数学家弗雷德里克·兰开斯特（Frederick Lanchester）提出了战争中的平方规律，即作战单元的数量以平方的形式影响战斗力，是比单元作战能力更重要的战争胜负决定因素。如果红蓝两方开战，蓝方作战单元的单元战斗力为红方的4倍，红方只需要在数量上集中2倍于蓝方的作战单元就可抵消蓝方在单元战斗力上的优势，这也可以看成是“人海战术”的科学原理。正因如此，无人机集群能够实现对少量优势战机更大的胜率。无人机集群将原本造价高昂的多任务系统分解为若干

低成本的小规模作战平台，将原本单一的功能平台分散到大量异构、异型的个体中去，系统的倍增效益将使无人机集群具备远超单一平台的作战能力。相比于技术高精尖但造价高昂的传统战机系统，无人机集群可以以更具成本效益的方式挫败对手。

人类，你**out**了
——机器人终将淘汰人类？

科技对社会的影响是一柄双刃剑：一方面让少数风口浪尖的弄潮儿越发举足轻重；另一方面则让更多身无长技的吃瓜群众无事可做。智能机器人的进化也不例外：当计算机变得足够聪明之后，一定会取代人类完成很多需要高级训练水平的工作。但我们大可不必杞人忧天，因为在两个多世纪前，这正是工业革命的直接结果。

拜工业革命和信息革命所赐，高速发展的自动化已经抢走了两个世纪前99%的工作岗位，效率低下的人力和畜力被先进科技带来的高级生产力所取代。但这些被取代的劳力并没有成为时代的弃儿，因为先进科技在全新的领域里又创造了数以百万计的就业机会。这些曾经面朝黄土背朝天的人们离开土地，涌入工厂，开始处理各种各样的工业产品，这让一波接着一波的新职业纷至沓来。电器修理工、胶印工、食品化学家、摄影师、网页设计师，这些新职业都建立在自动化的基础之上，这显然是两个世纪前我们的祖先无法想象的。

今天的情形像是19世纪的翻版：工业革命带来的先进科技曾经取代了人力和畜力，人工智能革命带来的进化版机器人也终将成为构成庞大工业体系的一颗颗螺钉与齿轮。想象一下在不久的将来可能出现的场景吧：在工厂里，这些敏捷的机器人能一整天地举起重达150磅的物品，对它们进行检索、分类与装载；在专业农场里，由机器人完成水果和蔬菜的采摘同样是不可阻挡的大势所趋；在药房里，分发药剂的工作将全部交给后台的机器人，药剂师的工作重点则在前台提供咨询服务。在办公室和学校里，可深夜工作的机器人将接管清洁工作，它们将从最简单

的拖地板和擦窗户做起，最终还能清洁厕所；在深夜的高速公路上，长途货车的驾驶室里坐着的也将是永不疲劳的机器人，根本不必担心疲劳驾驶造成的车祸。

也许有人会说：劳力者治于人，而劳心者治人。相比于简单的生产性活动，机器在管理规划方面的表现还达不到人类的水平。遗憾的是，现有的科技进展表明，这种情况也不会持续太久：软件能够直接通过体育比赛的统计数据撰写新闻报道，或从网络的只言片语中提取某个公司每日股价表现的概要。和文书处理有关的任何工作都将由机器人接管，即使是那些非文书的医学领域，如自动化的诊疗与手术，也正在日益机器人化。需要死记硬背的任何信息密集型工作也都可以自动化。机器人对人类工作的取代将如史诗般宏伟壮阔，而且不可阻挡。

有些工作人类能做，但机器人做得更好。人类可以编出棉布，自动编织机也可以，但自动编织机的效率更高，成本更低。相比之下，人造产品唯一的优势就是瑕疵——这不是嘲讽，而是真实的情感。瑕疵在告诉你这些产品出自于和我们一样具有喜怒哀乐的人类，而非冷冰冰的机器。在不久的将来，这些瑕疵必将演化为某种亲切感。可在其他容不得瑕疵出现的行业，机器人就会有无可比拟的优势。电脑抵押贷款评估正在大规模地取代人类估价师。税务准备、医学图像分析以及法律证据收集等行业的计算机化程度也在逐渐提高，可就在不久之前这些任务都依赖于高薪雇佣高智商的人群——想想影像科医生或是律师的薪酬吧。我们已彻底接受机器人在制造方面的可靠性，很快我们接受机器人在智能和服务方面的可靠性。

事实上，没有什么工作是人类不能完成的。只不过在分工日益精细化的今天，人类更重要的任务是操作机器，因而在不依赖机器情况下的工作能力的确在逐步退化。我们无法凭借自身的力量制作精密仪器的芯片，或是在网络中快速搜索出北京到广州的机票价格——这些工作所需的精度、控制性乃至高度注意力，是人类的身体所不具备的。我们不具

备如此的注意力，以在每个断层扫描图中的每个像素中寻找癌细胞的蛛丝马迹；我们也不具备如此的反应能力，将熔融玻璃膨胀成瓶子的形状；我们更不具备如此的记忆力，跟踪中超联赛十多年来的所有比赛结果以计算出正确的赔率。机器每一次取代人类的工作都会演变为轩然大波，但实质上，它们只是占有了我们无法做到的工作机会，同时也扩展了我们对宇宙的认识。

在机器人和人工智能的协助下，多少天方夜谭式的预言已经成为现实：通过肚脐切除内脏中的肿瘤；制作催人泪下婚礼的视频；在火星上遥控探测车；在衣服上打印出远方朋友通过无线电波传来的图案。如果生活在19世纪的人们能够穿越到今天，他们一定会眼花缭乱、目瞪口呆。这正是机器帮助我们实现的梦想。因此，面对智能机器人的汹汹来势，问题并不在于我们能够做什么，而在于我们应该做什么。

尽管大部分工作将由机器人完成，但这不意味着人类丧失存在的价值。作为人类的我们需要定义和发现新的任务，这些任务就是机器人的未来工作。无人驾驶汽车的普及将催生出行程优化师这个人类新工种，其使命就是对交通系统进行优化以节约能源和时间；常规的机器人手术亦将需要新的技能，以保持机器的无菌运行；当可穿戴设备记录下人体林林总总的数据后，专业分析师这个新职业将帮助你解读这些数据的真正意义……每个人都将拥有个人专属的机器人，但拥有并不意味着成功，成功只属于那些对机器人的工作流程进行组织、优化和定制创新的人。这就是人类与机器人的共生图景：我们人类的任务将是继续为机器人寻找工作，而且，这个任务将永无止境。

脆弱的三定
——奴隶、伙伴还是主人

从构词法中可以看出，机器人的本意是服务于人类的工具，人类与机器人之间是不可撼动的主从关系。但随着机器人智能的不断进步，人类与机器人之间的关系也在悄然改变，仿人机器人的出现就是这一改变的结果。

仿人机器人（Humanoid Robot）是与人类外观特征类似，具有一定的智能和灵活性，能够与人交流并不断适应环境的新型机器人。目前，最高水准的仿人机器人已经可以达到以假乱真的地步：在2005年爱知世博会上，大阪大学展出了一台名叫Re-pliee Q1expo的女性机器人。该机器人的外形复制自日本新闻女主播藤井雅子，动作细节与人极为相似。参观者很难在较短时间内发现这其实是一个机器人。十年后的2015年，大阪大学和京都大学联合研发出一款可以使用人工智能流畅对话的“美女”机器人Erica.Erica是位23岁的气质美女，相貌通过扫描多名人类美女的脸部特征合成，声音与真人发音非常相似。通过眼睛、嘴唇和脖子等多处部位的气压活动，Erica还能够惟妙惟肖地表现各种人类表情。2016年，中科院也发布了我国自主研制的仿人机器人“佳佳”，意味着我国首台特有体验交互机器人的正式问世。“佳佳”的外貌非常逼真，与真人等比例制作，五官精致自然，细节非常出色，并且初步具备人机对话理解、面部微表情、口型以及躯体动作匹配、大范围动态环境自主定位导航和云服务等功能。

图5-6 仿人机器人Erica

仿人机器人固然可以像工业机器人一样从事生产活动，但它更重要的意义是作为人类的伴侣而非工具出现，目的在于和人类达到良好的身体和情感互动，这也是为什么它们通常被制作成人类美女的样子。虽然以现在的技术水平，仿人机器人虽然还达不到心有灵犀的水准，但光是它们萌萌的外形就足以让你我心生怜爱了。而且不要忘了，作为宅男神器的高科技性爱机器人也已经诞生了呢！

可是，当我们在谈论仿人机器人时，我们到底在谈论什么？谈论的究竟是具有机器人形体的人，还是具有人类思维的机器人？

这样的问题本质上就是人类智能与人工智能的关系问题。如果有一天，人工智能进化到与人类智能同样的水平，机器人是否还会满足于对人类的从属地位？反过来，如果人工智能有可能在人类智能的基础上更进一步，机器人又会如何对待人类？而人类智能与人工智能又能否共生与融合，形成人机合一的终极永生模式？

图5-7 电影《摩登时代》海报

《摩登时代（Modern Times）》中令人不寒而栗的景象更像是机器人的终极归宿。由英国喜剧大师查理·卓别林（Charles Chaplin）自导自演的这部影片真实地刻画了社会底层的工人在资本压榨与利益驱使下的悲惨境遇。但在将来，这样的故事可能不再发生，因为我们已经有机器人来从事这些本就应该属于机器人的工作：它们不会出错，它们不知疲倦，它们不懂反抗。在当今社会严密的生产链条中，高效率、低成本、准确性成为衡量个体价值的标准，个性化的行为则称为无法控制的破坏性因素。机器人技术的发展则解决了这一难题：与其让人异化为机器，不如让机器淘汰人。

但这样的机器人归根到底不过是人类自身能力的延伸，以人类工具的形式出现。可是技术本身同样具有进化的动力，如果有一天，机器人的生理和心理都发生飞越式的变换，那么人类是否会反过来被机器人驱

使，还真是一个没有答案的问题。美军已经开始尝试使用一种多足机器人用来排雷，每次它用脚引爆一颗地雷就会丢掉一条腿，但还会重新爬起来用剩下的腿继续前进。它的第一次实弹测试就被负责执行的上校半途叫停了，这位上校无法忍受看着它用仅剩的最后一条腿拖着烧成焦黑、布满伤痕的残躯蹒跚前进，寻找下一颗地雷。上校说，这个测试是“不人道”的。

探讨人类智能与人工智能关系的科幻作品可谓汗牛充栋，这一主题也是以好莱坞为代表的电影产业中永不褪色的黄金主题，而斯蒂文·斯皮尔伯格的电影《人工智能（AI）》或许能带来一些启示：影片中的机器人大卫是一个有感情的孩子，它被一个没有孩子的家庭购买回去替代自己的孩子。与机器的钢筋铁骨不同，人类的家庭让大卫感受到温暖的关爱。在经历了沧海桑田之后，大卫开始寻找自己的生存价值：脱胎换骨成为真正的小孩，重新回到莫妮卡妈妈的身边……

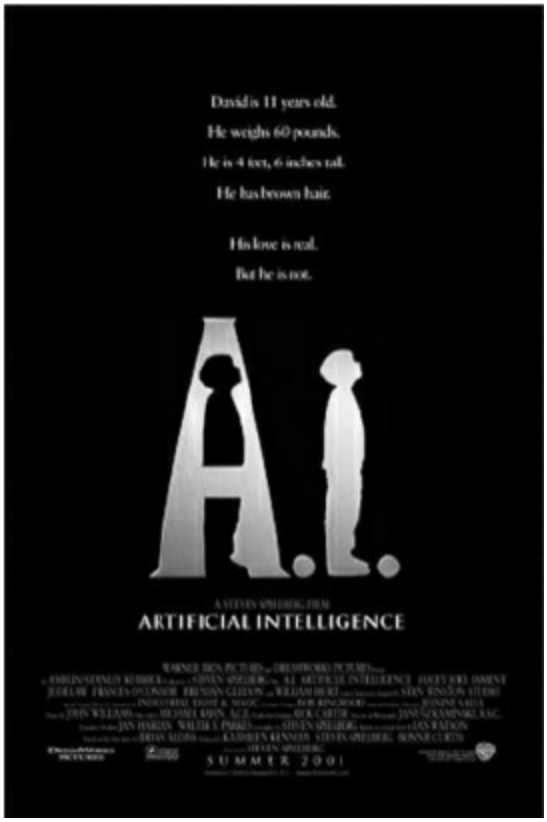

图5-8 电影《人工智能》海报

机器人的本义作为人类劳动的延伸产物，可随着人类将机器人做成了人的样子，人类与机器人之间的界限已经有了模糊的趋势，当机器人具备人类的智慧和情感之日，也就是人类与机器人之间的鸿沟彻底弥合之时。面对自己缔造出的他者智慧，理解、信任、尊重这些世界观价值观中的基本规则，甚至于“人”的概念，恐怕都要重新定义——人，不是生存状态，而是一种属性。人之所以为人，因其具有人性。

当影片中的机器人为人类服务时，你是否想起太平洋与大西洋上被血泪浸透的奴隶贸易？一艘艘贩奴船的朽木白骨之上，资本主义金碧辉煌的大厦拔地而起。当影片中的人类毫不掩饰对机器人赤裸裸的歧视时，你是否想起黑人小孩被荷枪实弹护送进白人学校的场景？手捧《独立宣言》的自由女神已不记得见证过多少对有色人种虔诚的诅咒。当影片中的机器人在屠宰场被随意毁灭时，你是否又想起人类历史一桩桩有组织的屠杀？自由、平等、博爱这些炫目的愿景已经在旁遮普、卢旺达、斯雷布列尼察的血海中酿成醉人的美酒。

我们如何对待机器人，就是我们如何对待自己。

至少在可以预见的近未来，机器人还只是人类身体的延伸，再强大的人工智能也还需要人类的思维来行使它的意志。这种灵肉的分开避免了很多道德困境。当有一天机器人也具备了人类的智慧，它们还会是工具和玩具吗？它们会如何看待自己的命运呢？人类又会如何看待它们的命运呢？也许那时它们就该称为“他们”了。

在人工智能的滚滚洪流中，只有一件事不可逆转：既然是智慧，就会存在对爱的渴求。

第六章　拒绝美丽新世界
——为什么我们还是人类

虽然名气稍逊于另外两部，但英国哲学家阿尔多斯·赫胥黎（Aldous Huxley）的作品《美丽新世界》确实是反乌托邦三部曲中最具深意的一部。赫胥黎关心的是科学进步对人类个体的影响：科技给人带来的改变过于剧烈，倘若人们不能控制好它，它就将要求人去适应它们，为它们所奴役。在人工智能托管一切的趋势下，相比于对机器超越人类的过度思考，人类异化为机器才是更为恐怖的事情。

图6-1 《美丽新世界》1931年初版封面

电车的困境
——道德代码如何编写

随着人工智能从科幻文学走进真实生活，道德准则的问题也不可避免地被提上日程。如果未来的人工智能仍然仅仅是纯粹意义上的非生物体，没有独立的意识、直觉与感情，也就谈不上要从道德、法律或是权利、义务等概念上对之作出限定；但反过来，如果人工智能本身具有机器智能、神经系统与生物体的融合性，具备了类似于人的意识、学习能力和感情，上述问题就变得无法逃避。随着无人驾驶技术的迅猛发展，眼下人工智能最著名的道德困境就是无人车版本的电车难题（The Trolley Problem）。

这个致命的伦理学实验的内容是：一个疯子把五个无辜的人绑在电车轨道上。一辆失控的电车朝他们驶来，并且片刻后就要碾压到他们。幸运的是，你可以拉一个拉杆，让电车开到另一条轨道上。然而问题在于，那个疯子在另一个电车轨道上也绑了一个人。考虑以上状况，你是否应拉杆？

电车难题是对伦理哲学中功利主义的批判。功利主义的哲学观点认为，大部分道德决策都是根据“为最多的人提供最大的利益”的原则做出的。当功利主义者面对电车问题时，会毫不犹豫地选择拉拉杆，用一个人的命换五个人的命。可批判功利主义的观点认为另一个人同样是无辜的，拉拉杆行为的实施者会成为一个不道德行为的同谋——为另一条轨道上单独的一个人的死负部分责任。如果囿于拉与不拉的困境而选择袖手旁观，这样的行为亦不可取——不作为将会是同等的不道德。

电车难题的核心在于，现实情况经常强迫一个人违背他自己的道德

准则，并且还不存在完全道德的做法。在无人驾驶领域，当面临车毁人亡和伤害其他路人的两难选择时，人工智能如何在危急时刻做出千钧一发、生死攸关的决定？在这样的场景下，机器人三定律似乎也失去了效力。根据报道，奔驰的下一代无人车将会被设定为任何情况下都拯救车上的乘客。从编程角度看，这属于一类算法派生，这种派生的主题是特权阶级在任何其他生命面前都有一条优先通行车道。这样做更接近于人类司机的本能，也避免了程序员代替人工智能“扮演上帝”的责任——眼下人工智能还没有自我意识，因而没办法自行作出决定，更没有办法对其追究责任。2016年夏，美国麻省理工学院的研究人员推出的“道德机器”网站提出的正是这类问题。在各种假设交通场景下，用户必须做出一系列的双输决定——也就是不管怎么决策都会有人死。任何人都可以进行“测试”，测试完之后网站就会公布你的“分数”以及与其他人所做道德决策的对比情况。

这类道德机器测试的可怕之处在于，人工智能做出道德决定的数据集中包含着更多的不确定参数：通常来讲，数字是权重最高的评价参数，机器可以简单地把拯救数多于杀死数，但年龄、性别、种族同样也可以成为不容忽视的因素，年龄是甚至会决定某些场景下的价值判断：如果无人车知道自己搭载的是一对耄耋之年的老夫妇，杀死他们以拯救一个年轻的三口之家可能会显得更有道德，甚至老夫妇自己可能也会做出这个无私的决定。一旦引入对社会的价值的概念，问题就会变得越发复杂——人工智能可以为了拯救一位诺贝尔奖级别的物理学家而牺牲整整一校车孩子的生命吗？这将引发空前的道德争论：君不见前些年大学生舍身救老农的社会事件造成的讨论有多激烈，但谁又能保证这一车孩子里不会出现另一个伟大人物呢？

把生死决策交给人工智能背后存在着强烈的反乌托邦隐喻，并如多米诺骨牌一般牵连出一系列复杂的相关问题。当偶然性、命运、天启这些心理缓冲被去掉后，对悲剧事故的追责无可避免地会引发一些伴生的问题，有些热爱自己动手的极客可能会玩弄系统，对自己的车辆进

行“越狱”，以便推翻有可能对自己不利的代码，确保自己成为事故的幸存者。突然之间，这把抱怨摆弄技术资产“权利”的意志论者推到了理论上“牺牲小我”观点的对立面。此外，人工智能的道德问题对无人驾驶汽车的私人产权也有决定性的影响。就理性而言，要把熙熙攘攘的鲜活生命交到人工智能手中，不同的无人驾驶系统就应该具备同样的技术水准，作为它们具备同样的道德水准的基础。据此而言，奔驰或特斯拉的模式只不过是公共系统或者高度管制化系统的垫脚石罢了。但是这样的结果会不会是制造商或者数据公司喜闻乐见的呢？

当然，有关人工智能的道德讨论不仅仅限于无人驾驶这一单一的应用场景。科幻与现实之间的界限越发模糊，但用于推动决策的价值观指向却鲜有提及。2016年9月，IBM、微软、谷歌等业界巨擘的计算机科学家联合成立了一个新联盟，名字叫做“造福社会与公众的人工智能合作伙伴关系”。该联盟的目标之一是建立人工智能领域最佳实践的规范标准，其中就包括了类似无人车电车难题这样的伦理问题的处理。

但在这些伦理问题解决之前，人工智能武器已经上路了。2016年3月3日，美国国防高级研究项目局透露，美军正在开发新一代智能型电子战系统。这套系统以深度学习技术为基础，能连续不断地感知、学习和适应敌方雷达，从而有效规避敌方雷达探测。在以往靠人力需要几周甚至几个月才能完成的敌方雷达搜索任务，在人工智能的帮助下可能几分钟就能搞定。这并非美国在人工智能武器领域的吃螃蟹之举，基于人工智能的更加致命的机器兵种早已开始在美军中服役：美国海军陆战队的部分基地已由移动式机器人值守；空军F-35战斗机的操作系统中包含几千万条代码，已经具备了部分自动控制的功能；海军X-47B无人轰炸机则更进一步，完全由电脑操控的它们代表了军用人工智能的最高水平。

图6-2 服役于英国空军的“掠食者”无人机系统

随着人工智能在军事领域应用的不断深化，一个重要的伦理决策渐渐浮出水面：支持还是反对基于人工智能的致命性自主武器系统（Lethal Autonomous Weapon System）。这些完全不需要人类干预的自动武器将自主决定将哪些对象作为袭击目标，袭击的强度又会达到何种程度。继火药和核武器之后，致命性自主武器系统可能第三次改写战争的面貌。

包括英国著名物理学家斯蒂芬·霍金（Stephen Hawking），特斯拉电动车发明人埃隆·马斯克（Elon Musk）和苹果公司联合创始人斯蒂夫·沃兹尼亚克（Steve Wozniak）在内的诸多科技界人士联合发表声明，要求联合国以国际法的形式禁止对自主武器系统的研发，但国际人道法律尚未对此类技术做出任何有针对性的规定。1949年签订的战争人道主义准则《日内瓦公约》要求任何袭击需要满足3个条件：军事必需、参与战争者和非参战者差别对待、军事目标价值及其潜在附带伤害的平衡。而当前的自动武器显然不具备作出这些主观判断的能力。国际社会对于此类武器也是反响不一，德国等许多国家迫切要求签署禁令，但美国、英国和以色列这三个在自主武器系统技术中领先的国家则认为这样的条

约完全多余。

确保人类对武器的控制已经成为道义上的当务之急。人类对人性的尊重会成为他们伤害他人时必要的抑制因素，相比之下，无生命的机器根本无法真正了解它们所伤害的生命意义何在，这是对人类尊严基本原则的侵犯。即便是在最残酷惨烈的战争中，无辜平民的安全也会远远重要于预期的军事利益，指挥官们会根据以往的经验和道德顾虑做出他们的判断并逐个确认，以免造成不必要的非军事人员伤亡。但指望“杀手机器人”理解这些概念似乎并不现实，隐藏着这些先进技术背后的，可能是空前的人道主义危机。

幸运的是，虽然人工智能尚不具备道德的判断力，但人类还有。物理学家们曾激烈反对在战争中使用核武器，生物学家也曾激烈反对在战争中使用生化武器，今天的人工智能和机器人领域的科学家们，已经担起了同样的道义重担。

电车的困境
——道德代码如何编写

与阿尔法狗对弈后，欧洲围棋冠军、职业二段樊麾和李世石都曾提到，阿尔法狗不同于人类，完全没有心态的波动变化。而即便是经历过大风大浪的李世石，也难免被心情的变化影响发挥。没有情绪、沉着冷静、不知疲倦，这是人工智能的机器本性。

在人工智能的发展过程中，有没有可能出现情感呢？这个问题可能并不像想象的那么简单，在人工智能先驱明斯基（参见第一章）眼里，情感甚至是智能的先决条件。他在著作《情感机器》做出了如下论断：“情感、直觉和情绪并不是与智能不同的东西，而只是另一种人类特有的思维方式。情感是先于理智存在的，人工智能只有智力，没有情感，不是真正的智能。”

从生物学的角度，情感的本质是高等生物对于外界环境的一种特殊反应机制，是复杂的主体意识判定客观事物是否符合自身需要的一种态度反馈。它的产生必定是受到客观世界的影响，和客观世界保持着某种关联。而它的机制则反映了主体对于这种客观影响的判断。比如婴儿见到母亲就会笑，听见打雷就会哭，这就是情感这一评判体系对于外界的影响所得出的判断。

人类之所以会产生情感，是进化的必然结果。为了适应复杂的生活环境，人类逐渐形成了发达的大脑。而发达的大脑可以产生超越其他生物的复杂情绪，反过来又可以让人更好地进行社会活动。在彼此之间相互作用下，人类的情感越来越复杂的同时，也对人类群体本身起着不可估量的作用。人类通过情感识别客观价值，通过情感交流表达价值，并

可以通过情感强度来计算价值。用学院派的话说：人类的情感，就是人类主体对于客观事物的价值关系的一种主观反映。

情感作为一种高级的信息处理机制，在大脑的学习和决策中的作用不容忽视，是智能不可分割的组成部分。情感细腻的人不会在别人气头上添油加醋，也不会在别人愤怒时火上浇油：他会感觉、识别并理解对方的状态，进而辨别分析什么才是恰当的反应，再将这样的反应恰如其分地表达出来。这显然是人工智能的短板：如果说人工智能的智商水平已经达到了金字塔的尖端，那么其情商绝不会超过一个五岁的孩子。在机器学习算法工具日渐普及的今天，人们猛然醒觉人工智能的发展离不开神经科学与心理学，下一个技术突破必将来源于对情感和情绪的了解，这无疑是现有人工智能的瓶颈。与此同时，今日的人工智能研究在各个方面似乎都因为忽视了情感、无法充分理解情感的各项机制，而难以取得进展。

图6-3 罗莎林德·皮卡德

如果人工智能可能发展出情感的话，其机制就必须建立在价值判断

的基础上，这一领域的研究被称为情感计算（Affective Copmuting）。情感计算的概念由麻省理工学院的罗莎林德·皮卡德（Rosalind Picard）提出，她也是这个冷门方向的主要研究者，正是她的著作《情感计算》开创了这个机器学习与神经科学的交叉学科。

所谓情感计算，意在通过赋予计算机识别、理解、表达和适应人的情感的能力，来建立和谐的人机环境，并使计算机具有更高的、更全面的智能。在人脑中，牵动情感的神经网络是一个复杂的整体，而情感本身也必然会参与大脑的决策过程。认知和情感不是井水不犯河水，而是你中有我我中有你。可目前的人工智能只能按照预先编写的逻辑算法处理输入而得到输出，并不会受这些信息的影响，更别提让这种影响反作用于决策了。

情感计算要参与到人工智能中，首先需要高等级的人机交互，让机器能够准确地识别人的情绪与心境。然而现有的人机交互还处于直截了当的低级水平。皮卡德的做法是利用计算机视觉和其他技术，训练计算机识别人类的情感表达。麻省理工学院的情感计算实验室通过给人工智能配上足够多的传感器来获取诸如语音、面部表情、手势、站姿、脉搏、皮肤电流等生物数据，训练人工智能的感知能力，再基于这些数据建立起人类情感和生理行为之间关系的理论模型。建立起理论模型后，人工智能就可以借助机器学习的算法来识别出情感状态。根据识别的结果再由算法进行反馈与修正，最终实现具备情感的人机交互。相关机器学习算法的技术细节，在此就不做赘述了。

情感计算更高远的目标是让人工智能通过语音和图像实现语义识别，从上下文的判断中识别用户情感，进而了解用户的真实意图和需求。这样的人工智能能够从多个维度识别并理解我们的情感，提供专属的个性化服务，使用户对“情感机器人”产生情感上的信任和依赖。

2014年6月，日本软银集团旗下的机器人控股子公司推出了新型情感机器人Pepper，并于2015年2月正式发售，首发的限量一千台颇受追

捧，1分钟内即告售罄。Pepper小朋友身高120厘米，体重28公斤，配备10.1英寸触摸屏，有着萌萌的整体人形设计，电池容量可让其持续工作超过十二个小时。作为情感机器人，Pepper的卖点在于它内置“情感引擎”，通过分析人类的面部表情、语音语调和讲话内容来“读懂”人类当前的情绪，从而灵活应对各种情况。更重要的是，小朋友自身也有情感特征：当Pepper被主人忽略时，它会模仿出一种孤独和难过的状态，当受到表扬和沐浴爱意时，它就会变得开心。Pepper并非传统意义上的家政机器人，正如软银集团总裁孙正义所说：“机械式的机器人今后还会有市场需求，但我们想制造出来的，是能够理解人心，能够与人沟通的机器人，我们认为这是非常有价值的事情。”

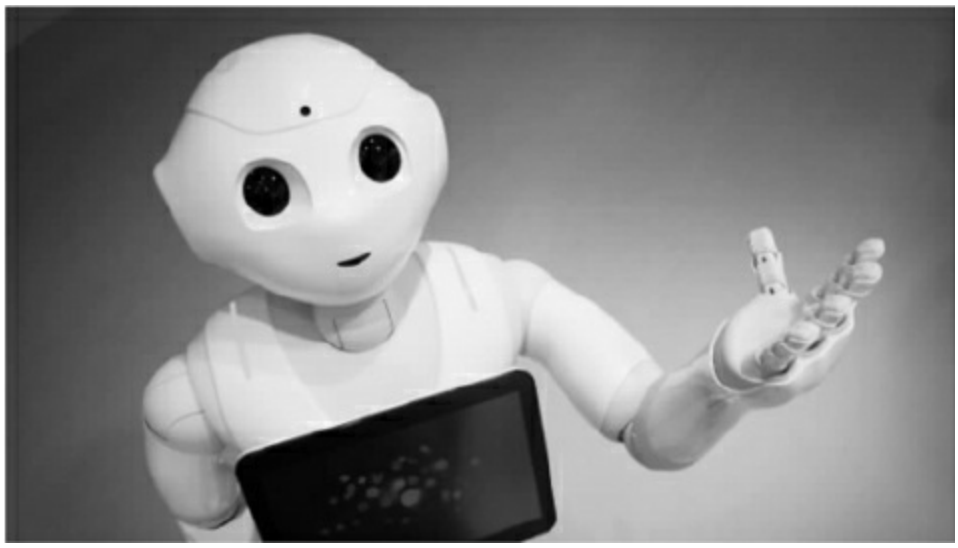

图6-4 情感机器人Pepper

如此这般的情感计算是否使人工智能真正具备了情感？这是争议的焦点所在。

这种通过机器学习算法结合海量数据训练得到的反馈还远远谈不上真正的情感。因为在处理这些传感信息的过程中，人类作为生物体会产生非常复杂而精密的化学反应，这样的反应远非数以万计的数学参数可以描述。众所周知，恐惧这种情绪是会传染的，当一个人感受到同伴的恐惧时，他自身也会受同样情绪的影响，并产生保护、战斗或者逃跑的

反应。也许最优秀的情感计算人工智能能够分析出对方的恐惧，并同样采取保护、战斗或者逃跑的行动，但其自身却绝不会真的感受到恐惧这种情绪。目前学界的主流观点认为：人工智能能够模拟喜怒哀乐这些直接的情绪，也能产生有这些情绪产生的反应，但诸如羡慕嫉妒恨之类的复杂情感尚不能明确地表示出来。

至于爱情这样复杂的生理与心理过程，还是把它作为人类的自留地好了。

创造性与想象力
——人类最后的阵地

在与李世石的第二场人机大战中，执黑的阿尔法狗的几手妙招让人刮目相看。

从开局起，阿尔法狗就不走寻常路，下出一步大概人类棋手永远也走不出的开局。而黑棋的第13手和第37手，是两手让内行外行的解说和舆论都大跌眼镜的两手棋。黑13手让李世石用去了一支烟的时间思考。古力九段表示“这两步，在人类的棋谱上没有出现过”，现场解说的麦克·雷蒙（Michael Redmond）九段则说“我不能确定这一手是好还是坏”，棋圣聂卫平更是认定这是一手值得“脱帽致敬”的妙手。待到黑棋下出第81手时，古力九段和俞斌九段的反应说是大惊失色也不为过：似乎颠覆了我们对围棋的认识。这局棋阿尔法狗看似赢得四平八稳，背后却是围棋界的惊涛骇浪。

阿尔法狗的棋风，让人想起围棋界一个久违的名字——吴清源。吴清源号称“棋圣”，曾在半个多世纪前的十番棋擂台上击败了当时日本所有的超一流高手。当吴清源横扫日本棋坛时，藏在恐怖胜率背后的是碾压型的优势：初盘看似愚蠢到家的棋路随着形势的推进到收官时都被证明是“神迹”。他对围棋的理解超越了整个时代，甚至现在也无人能及。吴清源对落子好坏的判断，对子力的效率的评估，对潜在关联的解读，都建立在高屋建瓴的判定和超乎常人的理解之上，似乎开启了上帝视角。

图6-5 吴清源

阿尔法狗在人机大战第二场中的表现，突破了人类对围棋的常规认识，隐隐透出吴清源的遗风。这局棋引起了阿尔法狗可能永远无法理解，但却对人工智能的发展意义重大的讨论——人工智能是否已经具备人类独有的创造力？

阿尔法狗诞生于大数据和深度学习，但用于学习的数据都是人类起手的对局。可正是在常见的对局中，阿尔法狗下出了从未出现过的招数。对此，计算机专家的解释是计算机现在的学习方式并不来源于简单的训练量的提升，而是在训练量的提升中不断修改函数联结与参数生成的方式。但由于神经网络是以黑箱形式呈现出来，作为创造者的我们也无法确知这些代码经历了怎样的反复纠结后，才让自己更上一层楼。可是，当画家或作曲家的灵感突然迸发之时，谁又能解释这是如何产生的呢？如果阿尔法狗的表现都不能被称为是具有创造力，那创造力又是什么呢？

用认知科学的话讲，创造力能够变具体为抽象，超越直觉、定势、

传统习惯的束缚，发现新的联系和关系，也就是个体在已有知识经验的基础上，利用多角度思维活动进而产生具有新颖独特特点和对社会具有使用价值的产品过程。而在狭义上，创造性思维是针对个体而言的，也就是产生对某一个体而言具有新颖独特特点和实用价值的任何思维过程。

对于人脑的创造力问题，心理学和哲学界还没有统一理论。研究表明，大量创造性思维和隐喻、类比推理有关。化学元素周期表的发现者德米特里·门捷列夫（Dmitri Ivanovich Mendeleev）借助扑克牌来探寻已知元素规律，做梦时突然“发现”元素周期表的故事，就包含感性隐喻和类比推理。计算机程序如果能模拟这种过程，就有可能具有更大的创造力。毕竟，如果人工智能想要有朝一日具备媲美人脑的创造力，就不能把创造力的来源用天启的灵感和悟性神秘化，而是应更深入地研究人脑创造性思维的机制，把它转化为一个图灵可计算的问题。

一位画家会反复训练自己的笔触，一遍一遍不厌其烦地绘制相同的东西；一位诗人也会反复地遣词造句，为选择一个最能达意的单字而思考不止。优秀的创意来源于日积月累的重复与尝试，古人所说“熟读唐诗三百首，不会做诗也会吟”也正是这样的道理。创造力来源于大量素材与技能的点滴拼凑，而非魔术般突然的闪现。从这个角度来说，有了足够规模的数据和足够强大的算法，让人工智能产生创造力也绝非缘木求鱼。

人工智能在可见范围内看不到具有基于欲望和激情的想象力和创造力的可能性，但它的的确确可以模拟这两者。几乎所有的艺术创作都高度依赖想象力和创造性，但其依赖性远远达不到我们观念中的高度。绝大部分小说都达不到《红楼梦》的艺术水准，而是可以从大量的小说素材中拼接出来。人工智能不太可能在现有领域之外开辟一种新的流派，但创作出具有起点中文网水平创造性和想象力的小说可不是什么难事。

2012年7月，伦敦交响乐团演奏了一曲名为《通向深渊

（Transitstoan Abyss）》的作品，有评论家认为它“充满艺术感并且让人愉悦”。可这首曲子并不是由作曲家谱写，而是由西班牙马拉加大学开发的名为“伊阿摩斯（IAMUS）”的人工智能算法在计算机集群上跑出来的，而且这个高产的“作曲家”已经创作了几百万首古典风格的作品。能够作诗的人工智能很可能永远都写不出《将进酒》《侠客行》这般作品，可问题是，纵观古今又有多少个李白呢？

真正的创造性与想象力来源于对生活深入的理解与细腻的感悟，艺术的最高境界也正是让观者与听者产生共鸣，心有戚戚。人工智能专家侯世达（Douglas Hofstadter）对机器的创造力有着极为精彩的论述：

“能有如此功能的‘程序’必须得能自己走进这个世界，在纷繁的生活中抗争，并每时每刻体验到来自生活的感受。它必须懂得暗夜里的凉风所带来的喜悦与孤独，懂得对于带来温暖爱抚的手掌的渴望，懂得遥远异地的不可企及，还要能体验到一个人死去后引起的心碎与升华。它必须明了放弃与厌世、悲伤与失望、决心与胜利、虔诚与敬畏。它里面得能把诸如希望与恐惧、苦恼与欢乐、宁静与不安等等相对立的情绪混合在一起。它的核心部分必须能体验优美感、幽默感、韵律感、惊讶感——当然，也包括能精妙地觉察到清新的作品中那魔幻般的魅力。”

这也是为什么我们仍然真诚地相信，创造性和想象力将会是人类最后的阵地。

不确定性的终结
——反乌托邦的终极奥义

反乌托邦（Dystopia）是一个内涵复杂的概念，从不同的角度可以得出不同的阐释。但从技术角度来说，人工智能是反乌托邦的基础——所有的反乌托邦社会都建立在高度发达的人工智能之上。发达的人工智能提高人类的生活水平，却让精神世界越来越虚弱空洞，这是一股不可逆转的潮流。

要从人工智能的角度理解反乌托邦，反乌托邦三部曲中的《我们》或许能给出最清晰的阐释。《我们》是前苏联作家叶夫根尼·扎米亚京（Yevgeny Ivanovich Zamyatin）于1921的作品，对其后面的两部反乌托邦作品都有深刻的影响。相比于乔治·奥威尔（George Orwell）笔笔如刀的反极权主义战斗檄文《1984》，《我们》的政治隐喻并没有那么强烈，却多了些对科技反制人类的隐忧，它的主人公正是一名数学家——也许这就是它为什么不像《1984》那么脍炙人口的原因。可正所谓无心插柳柳成荫，虽然并非出于本意，但扎米亚京阴差阳错地预言了苏联政治的发展走向，也就莫名其妙地变成了一个持异见者而漂泊他乡。

图6-6 叶夫根尼·扎米亚京

《我们》中的“我”是未来的大统一王国的数学家，也是宇宙飞船一统号的总设计师。在大一统王国中生活的人们是具有高度的一致性的，千万人如一体的“我们”，每个人没有姓名，只有编号——“我”是D-503.在王国里，日常生活按照《作息时间戒律表》有条不紊地严格进行，甚至连性爱也被进行了数字化地组织——在固定的时间凭票完成。大一统王国的科学进化得近乎完美。他们已经使用数学法则来创造诗歌（的确是高水平的人工智能！），还天才性地创造发明了“一致同意音节”，造就出最宏大的音乐史诗。

作为高级知识分子，D-503也会看看前人的古书，他会讶异于古人居然还生活在无组织无纪律的自由之中：“使我一直困惑不解的是：当时的国家政权怎么能允许人们生活中没有我们这样的守时戒律表，对用餐时间不作精确的安排，任人自由地起床、睡觉。有的史学家还谈到，当时的街上好像灯火彻夜通明，车马行人通宵穿行不息。”更难以置信的是“这个国家居然对性生活放任不管——这真是咄咄怪事：不管是谁，在什么时候，进行多少次，在什么地点……都由着人们自己，完全

不按科学规律行事，活像动物。他们也和动物一样，盲目随便地乱生娃娃，真让我觉得可笑！”

可是有一天D-503遇到了异性I-330，不由自主地从“我们”变成了“我”，并稀里糊涂地参与了一场推翻大一统王国的计划。最后，I-330被送进了一种叫作“气钟罩”的刑具里处死，D-503被捆在手术台上接受了切除幻想的手术，回归了与“我们”一致的生活。他在日记的最后一页坚定了对王国理想的信念："40号横街上已经筑起了一堵临时高压大墙。我希望胜利会属于我们。我不只是希望，我确信，胜利属于我们。因为理性必胜。”

反乌托邦的文学作品描绘出一幅幅恐怖的景象，可更恐怖的是，这是完全符合逻辑的合理结果。人工智能的最高层次就是不确定性的完全终结。

在《人在礉途之泰礉》里，成功人士徐朗与傻小子王宝之间有一段关于葱油饼的有趣对话：王宝的葱油饼一天能卖八百张，一个月净赚两万六千元。徐朗告诉他如果加盟费定在每月五千块的话，在北京开上一百家加盟店，一年就能拿到六百万。要是全国范围内开到五千家连锁店，就意味着每年可以赚三个亿的加盟费！看到商机的徐朗眼前一亮：

徐朗：你把配方卖给我吧。

王宝：行啊，我的配方就是，必须我亲自做，不能请人，不能速冻，必须得新鲜出炉。

徐朗：所以你一辈子只能做葱油饼，你知道吗！

王宝：你咋知道啊，我就喜欢做葱油饼！

工业革命已经终结了可计算过程的不确定性。大规模的工业化生产能够使经济腾飞，可流水线上生产出的产品千篇一律。相比之下，大部分奢侈品都秉承着手工定制的原则，这也使它们的身价达到了流水线产

品的千百倍。无论能工巧匠的技艺多么高超，都不可能达到机器的稳定性与精细度，可正是这些不可控制和不可克服的因素给他们手中的物品赋予了蓬勃的生命力：每一件都是独一无二的。魔幻现实主义作家加西亚·马尔克斯·加夫列尔（Garcia Marques Gabriel）错乱癫狂的时间与空间，先锋派作曲家伊戈尔·斯特拉文斯基（Igor Fedorovitch Stravinsky）斑驳陆离的调性与节奏，野兽派代表人物亨利·马蒂斯（Henry Matisse）直率粗放的色彩与构图，无不彰显着这狂野的生命力，这是再强大的算法穷尽所有数据也无法计算出来的生命力，因为科技本身形成的秩序足以碾压一切不规则的元素。

有时候我会怀疑，手机和电脑里的那个我甚至比我本身更像我自己：我的家里养的两只可爱的暹罗猫，每当我打开百度时，页面上便会布满关于猫咪的有趣推送，看了半个小时后，我已经忘记了原本要搜索的内容；每当我打开淘宝时，琳琅满目的指甲剪、化毛膏、零食罐头便会滚滚而来，信用卡的账单上就又多了一笔笔剁手的记录；每当我打开每日头条时，关于猫咪饲养与护理的种种讯息又让我觉得收获颇丰，我会把它们剪藏进印象笔记，却不知道是否还会再看。

图6-7 亨利·马蒂斯1905年的画作《戴帽子的女人》

在苹果音乐里，我建立了一个音乐频道，选择了自己喜欢的几支乐队和几首歌曲，它就会提供一系列其他的歌曲，还恰恰都是我喜欢的类型。睡前躺在床上，边听歌边逛亚马逊是我的习惯，在亚马逊买过几次书后，它开始贴心地提供它认为我感兴趣的书籍，这节约了大量选书的时间，我只需要在有折扣时下单付款就够了。每逢下雨的周末，我会在视频网站上输入“印度电影”碰碰运气，久而久之这几个字的输入都可以省掉了，网站甚至可以自己把北印度电影和南印度电影区分开来。

所有这些都是真实的我在二进制世界中的投影，但反过来，这些由算法和数据构成的人工智能却在重构我的生活。既然关于我的信息已经构成了一个全新的我。这个看得见摸得着的我是否还那么真实？人工智能比我更了解自己，它制定了二进制的我应该如何吃饭如何睡觉如何购物如何生活的种种规则，提醒真实世界的我不要做二进制的我不会去做的事情。人工智能会诱导人自愿放弃思考的权利，用种种潜在的锁链塑

造出一条看不见的流水线，用数据和算法生产出一个真实的“人”。

在数据主义的蓝图里，由超大规模数据和超强计算能力结合而成的外部力量将称为世界的主宰；当所有数据彼此相连时，这种力量将无远弗届，从而比任何事物更了解整个人类。在《1984》中，“老大哥”还有被推翻的可能，可这位“人工智能老大哥”已经在不知不觉中隐匿于无形，我们一手创造了它，却浑然不觉于它的存在。互联网批评家叶夫根尼·莫罗佐夫（Yevgeny Morozov）把这种现象称为“算法说服（Algorithmic Nudging）"。在他看来，算法说服的核心问题在于以产品工程设计的表象去掩盖社会工程设计的本质。正确的算法和海量的数据正在把独立思考的公民身份从人身上剥离出去，留下的只是痴迷于自我优化的人工智能的试验品。道德、情感、创造与想象——这些机器所不能做到的东西——决定了我们作为人类的身份。但在这个算法为王数据说话的时代：同质化倾向正在无可避免地越发严重，区分开来你我的特征正在消失。如果人工智能告诉我们，这个决策背后的选择架构是基于大数据，那么所有人都会不假思索地随波逐流。它如此事无巨细地影响着我的生活体验，会不会有一天变成由它来决定我喜欢什么。我想我们都希望能做智能的主人，而不是智能的奴隶。

在机械的乌托邦里，一千个人眼里只会有一个哈姆雷特。那时，一个D-503的代号也就足以囊括自身的所有价值。

第七章　黑镜照进现实
——警惕技术的反噬

科技是人类凝结了时间和精力的产物，然而在未来社会中却成为了一种异己的存在物，摆脱了人类的控制。在这一过程中，人类和科技之间呈现出了此消彼长的态势，即人类付出的时间和精力越多，形成的异化于他的科技就越强大，异化的程度就越严重，他受到科技的占有和奴役就越严重，而他自身的内部世界就越贫乏。在未来社会中，人类将成为科技的产品，而“人”的身份则退居其次。

《黑镜（Black Mirror）》是英国电视4台及美国Netflix公司出品的迷你电视剧，目前已经推出了三季和一集圣诞特辑。该剧分别以多个建构于现代科技背景的独立故事，表达了当代科技对人性的利用、重构与破坏。剧中的每个故事都可以从多个层面解读出不同的含义，本书则仅针对剧中的部分技术以及它们在剧情中发挥的作用，浅论先进的人工智能可能带来的灾难性后果。

图7-1 《黑镜》剧照

比你更像你自己
——当隐私成为奢求

《黑镜》第一季第三集《你的全部历史》中，存在着一个记忆永不遗忘的世界。通过在耳后植入记忆芯片，所有的记忆都可以以视频的形式保存下来，随时可以查看。

利亚姆是个律师，敏锐的职业第六感让他在一个派对上嗅到了妻子的秘密。反复观看派对上各色人等的言谈举止后，利亚姆越发确认妻子和其他男人有染。在他穷追不舍的逼问下，妻子向他坦白了那段不堪的往事——与前男友一个月的恋情。性情暴烈的利亚姆气愤于妻子对过往的隐瞒，更难以容忍这段记忆还存在在昔日男友的记忆里。暴怒的他痛殴了那个男人，强迫他把和自己妻子记忆全部抹去。

然而事情并没有到此为止。在这个男人删除记忆时，利亚姆意外地发现里面有一段18个月前的记忆——自己的妻子睡在床上的画面，那时他和妻子已经结婚。傻子都知道这是怎么回事。回家后，几近崩溃的利亚姆强迫妻子调出18个月前的记忆。目睹着妻子和那个男人18个月以前的性爱画面，麻木的他脑子里只有一个疑问挥之不去：自己到底是不是喜当爹？

酒醒之后，面对空无一人的房间，想起那些开心的往事，利亚姆拿起了剃刀……把植入的芯片取了出来……

在我国古代，宫廷中有专门记录皇帝日常行为举止的官员，这些记录结果被称为“起居注”。皇帝每天说的话做的事都会被记录在起居注中，作为后世修史的参考。更重要的是，在位的帝王是无法查阅关于针

对自己言论、起居的记载材料，更别提更改与删减对自己不利的内容了。正是我国历代史官构建了国家与民族的记忆传承，流传于齐太史简之间和晋董狐笔之下的浩然正气必将万古长存！

图7-2 清圣祖康熙起居注文本

今天的我们已经拥有了古代帝王才能享受的待遇，想想是不是有点儿小激动呢？

但今天的大数据和古代的起居录完全不同。如今已不需要专人记录记忆，互联网正在以无可比拟的全面性与持久性将所有记忆写成永久。数字化的革命使所有信息都可以被轻松复制，海量存储器的价格下降又大规模降低了数字信息的存储成本，数据库技术的完善使得存储的信息能够被轻易地搜索与调用，没有边界的网络又可以使这些信息被任意共享。这样的环环相扣减弱了我们对自身记忆的控制性。

更重要的是，所有的数据都由我们自身产生，但所有权却并不归属于我们。现在，几乎所有的网页在被打开时都会弹出关于缓存的说明：在一个不起眼的角落里看起来颇为弱势的对话框中，弹出的却是控制意味十足的霸王条款：本网站有存储并使用你浏览网站时产生的缓存信息的权利（或是表达同样意义但更加委婉的措辞）。同意之后你才能继续浏览网页，不同意呢？对不起，压根儿没有不同意这个选项。这不是询

问，这是告知。至于这些数据被拿来做什么，作为数据所有者的我们却一无所知。

曾几何时，无法抹去的记忆是克格勃或斯塔西的代名词：两者分别代表苏联国家安全委员会和民主德国国家安全部，其作用都是无所不在的内务监控：尽可能详尽记录每个公民的一切信息与行为。这两个机构曾因其鲜明的极权主义色彩而臭名昭著，直到如今还处在无穷无尽的口诛笔伐之中。可今天，我们却同样在舆论的影响下轻易地被技术征服，心甘情愿地把关于自己的一切交给计算机与服务器，甚至唯恐有半点遗漏。假如哪天在这些数据中诞生了足以毁灭人类的超级人工智能，他要做的第一件事就是嘲笑人类的愚蠢。

自我保护是人的本能，遗忘正是有效的自我保护机制。人类的记忆本来是特例，遗忘才是常态。遗忘在人类决策中扮演中心角色。它让我们能够及时行动；它让我们认识到过去的事件，却不受缚于它们。信息的共享意味着我们失去了对信息的控制，个人隐私的暴露达到前所未有的级别，随着网络传播力的提升，信息暴露导致的大范围传播与失控的局面已经成为常态。2008年轰动一时的艳照门事件与2014年的i Cloud艳照门事件已经演变成一场场群氓的盛宴，而这正是信息泛滥不可避免的后果。

2014年5月，西班牙人冈萨雷斯起诉了谷歌公司，原因是每当他在谷歌上搜索自己时，得到的首页首条都是“冈萨雷斯因陷入债务困难其房屋被拍卖”——要知道这已经是1998年的旧事了。冈萨雷斯认为谷歌搜出这件往事对他的名誉造成了损害，要求谷歌取消这篇报道的链接。谷歌公司的拒绝让这桩小事闹上了西班牙的法庭，可由于无先例可循，西班牙法官也理不顺谁对谁不对，又将这个案件转交到欧洲法庭。而欧洲法庭的裁决支持了冈萨雷斯的请求：欧盟最高法院裁定，允许用户从搜索引擎结果页面中删除自己的名字或相关历史事件，即所谓的“被遗忘的权”。根据该裁决，用户可以要求搜索引擎在搜索结果中隐藏特定

条目。随后，谷歌宣布已开始根据欧盟最高法院的裁定，在搜索结果中删除一些特定内容，给予用户“被遗忘权”。

相比于冈萨雷斯的案例，另外一桩事件则更加让人哭笑不得：一位化名“雷吉斯-V“的网民是一个正在求职的信息工程师，同时也是一位父亲，他曾经和一个相当不错的职位失之交臂。原因是什么呢？根据他的说法：“当面试官用谷歌进行了一番搜索后，我就在最后一刻失去了面试资格。当人们在谷歌搜索引擎上输入我的名字时，位于第一页顶端的几篇文章都来源于费加罗报（La Figaro）的网站，这些文章涉及到一起15年前审结的罪案。虽然文中出现了我的名字，但这压根儿就是个同名同姓的巧合啊！我费尽口舌向其他人解释这件事情，并提出自己没有任何犯罪记录，以证清白。可面试官已经产生了先入为主的怀疑情绪，并拒绝了我对这个职位的申请……”

欧盟裁决的被遗忘权让个体在自己的信息层次结构上拥有了一个小小的出口。

这一判例在实质上对这样一个问题进行了审查：搜索引擎仅仅是内容的托管方，还是个人数据的“监控者”？对于后一种情况，这个“监控者”事实上需要为呈现给用户的内容负责。对于这个问题，欧洲法院的法官给出的答案是谷歌在处理其服务器上的数据时扮演的是“监控者”的角色，因而应该为相关资料负责。他们还认为“搜索引擎的业务活动是对网页出版商业务活动的补充，既有可能对基本的隐私权产生重大影响，也有可能对个人数据起到保护作用”。

对个人隐私权利的保障与全面记忆带来的利益之间，冲突正在愈演愈烈。谷歌公司坚称欧盟的裁决称为“知识审查”——又一个臭名昭著且颇具蛊惑性的词汇。通过折中手段，谷歌把删除个人敏感链接的过程演变成夸张的闹剧，授意新闻媒体充当其愤怒的喉舌，因为后者唯恐天下不乱。但明眼人都明白，谷歌公司的说辞只不过是在遮掩这场辩论的真正细微差别。作为数据生产者的用户要求对数据的权利，作为数据消费

者的互联网巨头们则希望将数据完全据为己有。海量数据中蕴藏着社会演变的大趋势，在这种大趋势下，所有单独的个体都微不足道。

在大数据时代，记忆外部化的潮流已经不可阻挡，遗忘逐渐成为奢侈。更恐怖的是，在这个去中心化的巨网中，如利亚姆般取掉记忆芯片都会成为奢望。我们将会通过完美的记忆丧失一种人类的基本能力：在当下坚定地生活和行动。英国哲学家杰里米·边沁（Jeremy Bentham）指出：圆形监狱的设计使得囚徒从心理上感觉到自己始终处在被监视的状态，时时刻刻迫使自己循规蹈矩。这就实现了“自我监禁”。关于我们自身的数据散落在网络的各个角度之中，完整地构成了一座无形却更为残酷的数字化圆形监狱。

身陷囹圄的我们，却还悠悠然乐在其中。

不要相信眼睛
——虚拟现实的幻境

和平过后，人类又有了新的敌人：蟑螂。当然这个蟑螂不是小强，而是一种被病毒感染的奇怪生物：鬼一般苍白的脸、尖利的牙齿、低沉的嘶吼声。为了对付这些敌人，军方开发出MASS系统并植入军人的大脑。根据官方声明，MASS系统能向士兵实时显示一切情报：从敌人信息、到地形地貌，从战场结构到与无人机对接，这帮助士兵提高了消灭蟑螂的效率。

“大头兵”是个刚刚入伍的小伙，身手不凡的他第一次参加实战，就击毙了一只蟑螂，另一只蟑螂试图反抗也被他数刀捅死。听着战友的赞许，“大头兵”不禁心生得意。可在打扫战场时，他的眼睛被奇怪的绿光扫过，这让他时常出现头痛，视觉也会出现干涉卡顿。军医和军方心理咨询师认为这是首次作战后出现的不适反应。可事情的发展显然不符他们的预期，“大头兵”的情况越发恶化，直到一次任务中，他惊讶地发现所谓的“蟑螂”竟然是和他一样的人类……

这是《黑镜》第三季第五集《战火英雄》中的情节，这些情节不是毫无依据的假想，而是有据可查的空穴来风。随着美国天下布武政策而来的是大量退伍美军出现心理问题。美国国防高级研究计划局为了解决退伍军人在战场上留下的创伤后应激障碍问题，斥资数千万美元研究脑机接口程序，试图通过植入式脑机接口调节情绪，提高记忆力，甚至消除不愉快的战争记忆。这可以视为虚拟现实技术在战争中的首次应用。

正所谓艺术源于生活而高于生活，《黑镜》把这种虚拟现实技术从退伍后提到了入伍前，直接给士兵生成了辅助作战的幻象。这个想法也

来源于真实世界，美国国防高级研究计划局也在开展另外一项研究——神经工程系统设计。这类装置向大脑反馈电子听觉或视觉信息来弥补听力或视力缺陷，而体验质量将远高于现有技术所能达到的水平。至于把听觉信息和视觉信息弥补成什么样子，就看战场上有什么需要了。

……军方心理咨询师向“大头兵”解释了一切：在战争中很多士兵会不忍开枪，在战争后也会出现心理创伤。MASS系统的作用就是让士兵把敌人看成怪物，把惨叫变成嘶吼，这样就能消除士兵的道德负担，提高杀敌效率。“蟑螂”是个谎言，所谓“蟑螂”只是基因有缺陷的人类。为了人类的繁衍，这些具有高致病基因的人就必须被清除。

获知真相的“大头兵”暴怒不已，他想要从自己的大脑中删除MASS系统。可一旦MASS系统关闭，他的眼前就会闪回出两个无辜平民血肉模糊的惨状，这惨状正是他一手为之。不堪道德重负的“大头兵”只能选择重新投入战斗。当他复员时，亲密的爱人与他深情相拥，可在MASS系统之外，他的面前只是一幢空无一人的破败小屋。

在这一集里，《黑镜》提出了一个严肃的问题：有了虚拟现实和增强现实技术之后，我们是否还能相信“眼见为实”？

事实上，广义上的虚拟现实离我们并不是那么遥远，互联网和社交媒体早已构建出与真实生活平行的线上生活。在网络上，每个人都可以化身为自己想要成为的人物：向你言之凿凿揭示特朗普与希拉里选战内幕的政论专家可能是楼下的的哥老张，津津有味大谈中国外汇储备与未来楼市走向之间的关系的经济大咖可能是公司前台的保安小王，洞悉王宝强与马蓉从捉奸在床到劳燕分飞全过程中一切细节的传媒牛人可能是菜市场的李家媳妇，长篇大论给你讲解皇家马德里与巴塞罗那之间恩怨情仇的足球大师可能是刚刚开学的大二学生小赵……然而，这些人都有另外一个共同的名字——键盘侠。1993年7月5日的《纽约客（New Yorker）》上配发的题为“在互联网上，没人知道你是条狗！”的漫画，成为对网络空间虚拟性的绝妙嘲讽。网络是一个独立于现实的平行世

界，在这里可以以崭新的身份享受完全不同的生活。但对于键盘侠来说，他们的武器——键盘——却在时时刻刻提醒他们网络与现实的差距。

虚拟现实的出现则在网络的基础上更上一层楼：以沉浸式的方式进一步强化虚拟空间的真实感。虚拟现实和增强现实真正的革命性就在于它改变了人类与自然交互的媒介。一直以来，我们认知世界的方式是通过视觉、听觉、嗅觉、触觉等各种感觉，也就直观地认为各种感觉触及到的都是真实存在的世界。在虚拟和现实之间，曾经横亘着不可逾越的鸿沟。可随着科学技术的发展与进步，这条鸿沟已经变成了几乎看不见的细线。但虚拟现实的出现彻底颠覆了这种认识：感官和神经识别出来的只是外境，可外境和真实世界之间，却可能横亘一道鸿沟。虚拟现实通过欺骗我们的大脑，进而控制我们的意识。一旦进入这个虚拟的世界，你就会身不由己地被它操纵。

图7-3 “互联网上没人知道你是条狗”

在古今哲学家的思辨中，早有对虚拟现实的思考：两千多年前，战国思想家庄周在《齐物论》中提出了梦蝶的问题：过去庄周梦见自己变

成了蝴蝶，生动逼真的蝴蝶，这正是我想要的，那时已经不知道自己原本是庄周了。可突然间醒过来，惊惶不定间方知原来是我庄周。不知是庄周梦中变成蝴蝶呢，还是蝴蝶梦中变成庄周呢？庄周与蝴蝶必定是有区别的。这就可叫作物我的交合与变化。

法国哲学家勒内·笛卡尔（Rene Descartes）也意识到了真实的相对性："一切迄今我以为最接近于'真实'的东西都来自感觉和对感觉的传达。"如果感觉本身就是欺骗性的，那么世界又何谈真实呢："我怎么能否认这两只手和这个身体是属于我的呢，除非也许是我和那些疯子相比？那些疯子的大脑让胆汁的黑气扰乱和遮蔽得那么厉害，以致他们尽管很穷却经常以为自己是国王；尽管是一丝不挂，却经常以为自己穿红戴金；或者他们幻想自己是盆子、罐子，或者他们的身子是玻璃的。但是，怎么啦，那是一些疯子，如果我也和他们相比，那么我的荒诞程度也将不会小于他们了。"

由古至今，人类一直在追求"真实"的道路上前仆后继，无论何种流派都无法回避这样的事实：我们对于真实的认知建立在人类感官的基础上，即便纯粹抽象理念上的推演，也无法脱离大脑这一生理结构本身的局限性。当我们可以借助虚拟现实创造外部世界对人类感官的刺激信号时，就创造出一个等效的"真实世界"。而在这样的世界里，人类变成了制定规则的上帝，所有伴随人类进化历程中的既定经验与认知沉淀将遭受颠覆性的挑战。我们将重新认知自我，重新认识世界，最重要的是，重新定义真实。

图7-4 虚拟现实设备Oculus Rift

随着技术的进步，今天的虚拟现实将不可避免地扮演昨天的网络所扮演的角色：逃避主义者的避难所。出生在21世纪的孩子已经习惯了被网络包围的世界：滑动手机解锁可能是他们人生中最早（抑或是最糟？）的几个动作之一——网络就是他们的生活，数字化就是他们的生存方式——长大的他们必然会将真实的现实世界和虚拟的网络世界无缝衔接。同理，今天的我们相信：一个生长在现实世界的人拥有着正常认知教育的人，都无法忍受永远地生活在虚拟现实世界中。可当未来的孩子刚一出生便处在虚拟现实之中，并以这个角度来观察世界时，真实的世界在他们眼中，会不会将是无法容忍的现实呢？

虚拟现实的出现，恐怕会重新定义物质与意识之间的关系：你经历的是真实，还是你想要的是真实？

云端的永生
——思维克隆人的背后

《黑镜》第三季第四集《圣朱尼佩洛》出人意料地讲述了一个温情脉脉的故事。

未来的医疗系统运用最新的意识存储再生技术研发出新型治疗方法，圣朱尼佩洛这座虚拟怀旧城市因此诞生。在这座虚拟城市里，因车祸成为植物人数十年的约克夏找回了自己的少女时代，并与迷人的凯莉缠绵悱恻。可这段同性恋情在凯莉眼里这不过是过眼烟云。凯莉的生活中曾有过一场幸福的婚姻，可年轻女儿的意外身亡，让白发人送黑发人的痛苦纠缠了凯莉和她的丈夫多年。两年前凯莉的丈夫离世，他为了陪伴女儿，放弃了使用圣朱尼佩洛的权利。

在坟墓里长眠不醒，还是在圣朱尼佩洛享受永生？凯莉在两种选择中纠结不已。最终凯莉做出了决定：她接受了安乐死，身体在坟墓里长伴家人，意识则在圣朱尼佩洛永随约克夏。此时，约克夏和凯莉化为了意识芯片上高低错落的电流，圣朱尼佩洛服务器德接口上灯光闪动，意味着她们在遥远的云端永远相知相守。

《圣朱尼佩洛》是《黑镜》中鲜有的大团圆结局，可掀开唯美百合恋情的面纱，看到的仍然是个非常严肃的哲学思辨：如何看待虚拟世界里意识不灭的永生方式？

片中对永生实现的技术设定合乎情理：大脑的所有信息都被提取并上传到云端，因此对于虚拟环境的所有感知和认知依然是建立在现世感知认知的基础上。这恐怕也要归功于技术的进化：在第一季的第三集

里，还只是把所有记忆存储在芯片之内。2013年9月19日，身患肌萎缩性脊髓侧索硬化症的著名理论物理学家史蒂芬·霍金（Stephen Hawking）就曾在展现他生平的纪录片首映式上提出：“我认为思维就是储存在大脑中的一段程序，就像电脑一样，所以理论上我们是可以将大脑复制到电脑里面，提供一种在死亡后的生存方式的。”

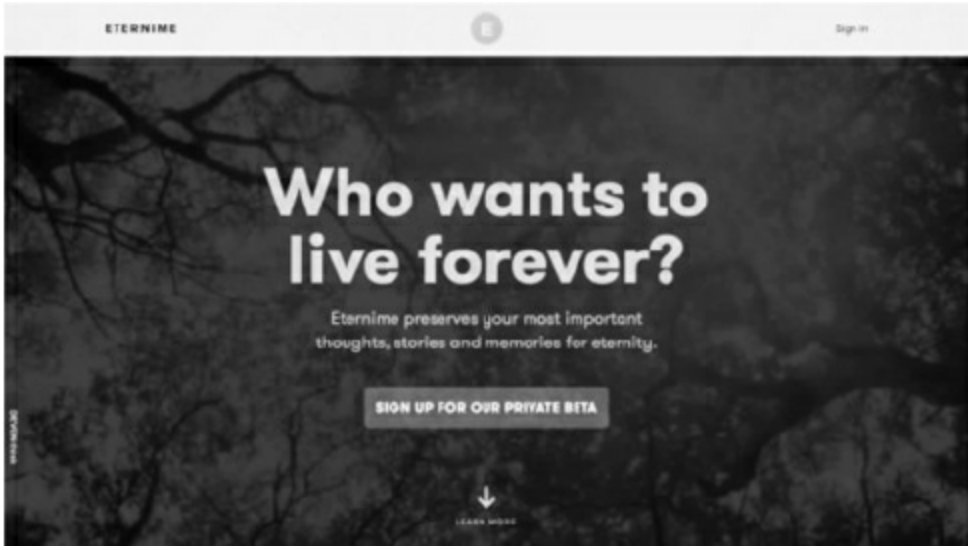

图7-5 思维备份网站eterni.me

这并不全是臆想：网站eterni.me正在提供这种服务。如果你在生前授权该网站访问你在互联网上的信息——电子邮件、脸书与推特、上传的照片、旅行的轨迹、通过谷歌眼镜观察并记录的事物以及其他，网站就可以基于这些信息生成你的人工智能化身，化身会尝试模拟你的外表和个性。如果你平时还与化身交流，它就会更了解你，也更能反映出你的特征性格。

思维克隆人概念的核心在于把躯体和意识剥离开来，把人的永生转化为人的意识的永生。这种简化可以说合情合理：德国哲学家卡尔·马克思（Karl Marx）在《关于费尔巴哈的提纲》中提出的极富洞见的论断：人是社会关系的总和，表达的也正是这样的含义。社会关系本身是看不见摸不着的，但它却塑造着一个人的思维方式，进而不断影响其命运。人工智能的发展使这一定性的论断得到了定量的扩展：人是由所有

意志和记忆构成的思维意识，躯体只是意识的载体。拥有和你一样的思维意识的就是你，而生物学躯体只是意识的一种承载方式；随着人工智能技术的逐渐演进，计算机存放意识的那一天恐怕为期不远。

人类意识是个包括知觉、感觉、察觉、潜意识等诸多方面的庞大复杂且难以描述的大集合，即便是通过图灵测试也难以真正模拟出类人的意识与思维。但大数据时代的到来为人类意识的重构提供了全新的思路：通过信息技术基质实现意识。互联网浏览历史、商品购买偏好、社交网络的留言评论、物理行踪的记录……信息时代的人已经被分解为数据化的自我意识。互联网公司Twitter已经推出了名为Lives On的软件系统，用来学习用户生前发布的内容，并在其死后以其风格继续发布。事实上，我们一直再以另外的方式做这件事情：所有的历史小说不都是在根据已有史料推断某个人在某个事件下的反应吗？可如果有一天软件系统达到了人类思考的复杂度，我们也就无法从判断秦皇汉武是否死而复生。

站在哲学角度，我们可以认为这个拥有相同记忆和行为偏好的第二意识，是独立于第一意识的存在。但是现代神经学的研究成果表明，大脑在人机合一的认知中表现出来的灵活性让人难以置信。一个截肢患者在安装假肢后不久，大脑就会为操控假肢而连接起更多神经元，将其作为身体的自然延展。据此我们也没有理由不相信，大脑也许能轻松地将成熟的思维克隆人接纳为“自我”的一部分，并延伸出生物大脑与软件大脑的并行意识。这种科技改变生活的体验，和现在的孩子从不知手机为何物到智能手机不离手的转变并无二致。

尽管听上去仍是那么科幻，但生物躯体的死亡不再意味着生命的终点，越来越多的科学家、程序员、商业家正在为此努力。人类此前从未体验过躯体之外的身份，但他们很快会体验到，永生时代也会随之来临。永生情结可能来源于对死亡的恐惧，但更重要的是对当下的眷恋。死亡的存在让我们明白生命的可贵，让我们学会珍惜现在，让我们明白

时间是自然给人最宝贵的礼物。总有一天会出现真正的圣朱尼佩洛，当日渐腐朽的身体无法承担仍然想恣意生活的灵魂时，它可以让人的思维挣脱时间的束缚，脱离肉体的桎梏在系统中永生。

可是在虚拟的永生中，一切都不会随着时间的流逝而失去，永生的价值又在哪里呢？或许对人类而言这不过是一种心灵慰藉：在某种程度上我们不会无迹可寻，借此获得某种超越物理湮灭形式的不朽。动物会通过交配与繁衍将基因这根接力棒代代相传，千万年生生不息之间保证了物种的永生。可对于人类来说，存在的意义不仅在于作为物种的人类，更在于作为你我的个体：从小到大拍下的照片、读过的书、说过的话，刻画出来的是各不相同的气质与性格——这些形而上的精神特质才是我们想要延续的，关于自己的记忆。

可即使人类意识活动的内容和心理轮廓能够被以数字化的方式记录下来，它所带来的影响的复杂和深远的程度也将远远超出我们的想象。这必然会给传统观念带来强烈的冲击，习惯了血肉之躯的我们怎会与机器人，甚至连机器人的外形都不具备的一团代码划上等号？更深层次的问题在于：“人”的定义到底是什么？这样的模拟大脑能否被视为“人”？如果被视为“人”的话，现有的社会又该做出怎样的改变以适应这些新人类？

另一个不得不思考的问题是：这样的“人”到底是为谁而活？离世的人希望被活着的人铭记，可活着的人终究也有告别人间的一天。当一个庞大的家族以思维克隆人的方式在服务器里重新汇聚一堂时，存在的只有慢慢褪色的往事，却没有任何新鲜的思维。约克夏和凯莉是幸运的，她们可以在无穷的岁月里相知相守，有永恒的爱情陪伴他们。可在圣朱尼佩洛，除了约克夏和凯莉相识相知的酒吧，还有一个地方叫做泥潭——在这里，行尸走肉般的永生人们醉生梦死，以扭曲的刺激寻找生活的意义。永生的他们也许腻烦了圣朱尼佩洛的生活，却没有办法逃离这个牢笼，只能如西西弗斯般忍受这无情的轮回，连选择毁灭的权利都没

有。这样的设定迫使我们思考思维克隆人冷峻的一面：所谓永生，到底意味着什么？

当死亡消逝之时，生命也就失去了意义。

无为有处有还无
——数据的黑洞

这是一个打分的时代，你的分数关乎你的命运。

当你和任何人产生任何社交行为后，都需要通过手中的终端系统为对方评分：1分代表你觉得他/她是个人渣，5分表示你认为他/她太赞了。所有分数的均值综合出关于每个人的结果——它不仅仅是个数字，更是社会地位的体现：4.5分以上的高分人群可以优先享受更高级的社会资源，比如租房福利、机场快速通道、更优质的教育和医疗服务等等；2.5分以下的低分人群即使健康健全也会被看作异类，失去工作和住房，被整个社会体制抛弃。

蕾茜的评分是4.2分，这显然不是个理想的分数，因为她看中的一处高档房产只对4.5分以上的高分人群才有优惠的价格，这让她烦恼不已。蕾茜的评分分析师一针见血地告诉她，分数低是因为她的社交圈子质量不高，这也决定了她很难在短时间内将分数提升到4.5分以上。要想快速取得高分只有一个途径：想方设法攀上高分人群的高枝。

蕾茜绞尽脑汁讨好身边的高分人士，奈何高分人士根本不吃这套。好在她还有位发小娜奥米评分达到了4.7分，为了引起娜奥米的注意，蕾茜在社交网上发布了一张和娜奥米一起制作的手工娃娃的照片。念及旧相识的娜奥米很快为这张照片打了5分，邀请蕾茜来做自己的伴娘，并在婚礼上发表演讲。获知参加婚礼的嘉宾都是4.5分以上的高分人士后，兴奋不已的蕾茜精心准备了关于儿时友情的演讲，希望能借此获得高分人士的好感，让自己的评分上升，成功入住梦想中的社区。

可蕾茜没有想到，等待她的是一场灾难……

这是《黑镜》第三季第一集《急转直下》中的情节，其深刻的内涵读者在观剧后自会体会，本书则只讨论一个小问题：对书籍、电影、歌曲的评价都可以以量化的分数形式体现出来，可这种方法真的能够套用在人身上吗？

美国哲学家丽贝卡·格尔斯坦（Rebecca Goldstein）在《柏拉图在谷歌总部》一书中，想象了古希腊哲学家柏拉图（Plato）穿越到今天美国加州谷歌总部的场景。在这里，一位软件工程师向柏拉图兜售他开发的“道德答案搜索引擎”，并和柏拉图争论起一个问题：什么样的生活最值得过？

在老派的柏拉图眼里，最值得过的生活无疑是热爱智慧的生活，也就是哲学家的生活。但谷歌工程师显然不这么认为：在他眼里，没有人比数据更真实。他的“道德答案搜索引擎”正是基于大数据的解决方案：收集所有关于美好人生的数据，然后根据你的个性偏好，为你量身定制一个最好的人生。正所谓“幸福的生活是相似的，不幸的生活各有各的不幸”。

不知不觉间，“量化自我”已经成为大数据时代技术宅们新的圣经。在这场如火如荼的量化盛宴中，各大硬件软件公司适时而动，以各式各样的可穿戴设备、应用程序和社交网络加入狂欢：手环会告诉你你平均每天行走的步数在三千至四千之间，行走距离为三公里左右，燃烧掉约一千二百卡路里；睡眠监测器会告诉你你一天的睡眠时间八小时，其中浅睡六小时、深睡两小时，早晨6:07到6:51之间睡眠质量最好，所以闹钟最好别设在6:30；谷歌眼镜会告诉你在过去的二十秒中，你的生气程度/高兴程度/沮丧程度/惊讶程度分别是多少；智能手机甚至会告诉你你的性能力有多强：软件厂商Ardenturous LLC开发出一款名为Spreadsheets的程序，它就能跟踪你做爱时的声音和时间，由此计算你的性能力——当然现在已经下架了。

图7-6 应用程序Spreadsheets界面

所谓“量化自我”，就是通过科技方式将自己日常生活的各方面，包括物质摄入、身体状况以及体能情况记录下来的一项活动。这项活动正在悄然演变为普通人的生活方式——你的手机里有没有这样的应用程序呢？人们都希望成为关于自己身体的科学家，一方面以自我认知、健康需求、结构化以及高效为由将生活的所有数据量化并记录下来，一方面则坚定不移地抱有这些数据必将有助于更好地管理自己的美好信念。

可如果拨开喧嚣的迷雾，“量化自我”更像是被炒热的噱头：我们身上到底有多少指标可供量化？如今的可穿戴设备和应用程序无非是在围绕身高体重、脉搏心率、行走距离、睡眠时间这些能够被体外监测的基础指标炒冷饭，实时监测血压血糖这些对身体健康更具意义的指标看起来仍然遥遥无期——没有哪个可穿戴设备的生产商敢于去冒从用户身上采血这样的风险。目前，营养摄入水平也进入了“量化自我”的量化范围：软件会根据你的饮食记录告诉你今天摄入了多少饱和与不饱和脂肪、多少必需和非必需氨基酸、多少B_1/B_2/C/D/E维生素、多少钾/钙/镁/锰微量元素……可愿意相信“量化自我”神奇魔力的极客们会相信这些数据吗？对这些数据准确率的量化恐怕是更具挑战性的任务。这些并不精确的数据更大的价值似乎在于维护量化者们掩耳盗铃的满足感。

在健康管理的实用性目的背后，“量化自我”还隐藏着一个神圣的精

神性假设：这些数据是检视自我的工具，它们揭示日常生活中的规律与相关性，帮助我们更清楚地认识自己。“量化自我”的终极目标正是发掘出我们是谁，我们想要什么，以及我们往何处去。然而事与愿违，在无尽的个人数据生产、处理与分享过程中，产生的只是勇夺朋友圈冠军时那庄严神圣的仪式感。

《福布斯（Forbes）》杂志的记者希尔曾就“量化自我”的主题大量撰文，在一次自我跟踪实验中她发现：“我最幸福的时候是在酒吧喝酒（啊）；最不幸的时候是在飞机上、在工作中（啊哈）；星期天是我最开心的一天，其次是星期三；我一个人时和跟人在一起时一样开心；我跟前男友互动比跟现男友互动更开心。”可这些数据又有什么意义呢？“该拿这些结果怎么办，我有点糊涂了。这意味着我应该工作时多去酒吧里待着，优化我的幸福感吗？还有，我应该重新考虑跟男朋友的关系吗？”

这也是个有趣的话题：婚姻的选择权由中世纪的牧师递交给了人文主义时代的“内心选择”——但在未来，接棒者也许是谷歌搜索引擎，“我会问它，约翰和保罗都在追求我，他们两个我都喜欢，但两种感觉不一样。我真的很难理清楚。鉴于你知道的所有信息，建议我怎么做？而谷歌会回答：我从你出生那天就了解你了。我阅读了你所有的邮件，记录了你所有的电话，知道你最喜欢的电影，你的DNA，以及心脏的所有历史生物信息。我有你每次约会的精确数据，我可以向你展示你每次与约翰或保罗约会时的心率、血压和血糖水平的实时图表。而且我了解他们就像我了解你一样。基于这些信息和我的高级算法以及几十年来关于数百万对恋爱关系的统计数据，我建议你和约翰在一起，因为从长期来看，有87%的可能你和他在一起更满意。”

庸俗的唯数据论一直都不受聪明人的待见。德国哲学家尼采（Friedrich Nietzsche）以他特有的语气发出尖锐的讽刺：“什么？难道我们真的希望把‘存在’贬低到像这样的地步：只要数学家在室内用一台

计算器就能‘运算’出来？更重要的是，人们不应该以为自己能理解存在的丰富意味——它需要良好的品味，先生们，那是一种敬畏感，对自己视野尚未看到的一切的畏感。”

数据的意义在于价值而非数字。如果人所有的行为、所有的感情、所有经历的事情、所有认识的人都可以进行量化，变成一组数据，那人类存在的价值在哪里？“量化自我”那迷人的面纱下面藏着的不过是一团虚无，三分之一的可穿戴设备会在半年之内被用户放弃就是再好不过的证明。

第八章　畅享技术红利
——人工智能的应用

有人担心这些机器人会夺走工人的工作。是的，机器人可以帮人们完成某些任务，任何一份工作都需要完成许多任务，而人工智能和机器人所创造的任务数量要比他们拿走的任务数量多很多。

对于那些追求生产效率的工作，机器人最合适不过，反过来说，人类对于那些追求效率的工作并不重要。不追求效率的工作也有很多，创新这件事情本身就是不追求效率的，因为在创新的过程中，必然会有很多失败；科学工作也是不追求效率的，因为实验室的工作就是不断试错的过程，如果不犯错误，就不可能学到东西；艺术也是不追求效率的；人际关系也是不追求效率的。

很多我们认为重要的事情都是不追求效率的，这些就是需要人适合去完成的工作，也是需要我们付出更高代价得到的东西。所以，追求效率的工作归机器人，不追求效率的工作让人去做。创造、探索、实验的工作，人可以比人工智能做得更好。

“人工智能”术语的创始人约翰·麦卡锡曾说过：“一旦一样东西用人工智能实现了，人们就不再叫它人工智能了。”可事实却是，在几乎所有需要进行思考的领域，人工智能都已经超越了人类。人工智能已经渗透到生活的方方面面，习以为常的我们甚至感受不到它的存在，只把它带来的种种便利视作理所应当。

通向巴别塔之路
——机器翻译

根据《圣经·旧约·创世纪》中的记载，大洪水劫后，诺亚的子孙们在巴比伦附近的示拿地定居。说着同样语言的人类联合起来兴建希望塔顶通天能传扬己名的巴别塔，这让上帝深为人类的虚荣和傲慢而震怒，更不能容忍人类冒犯他的尊严：如果人类真的修成宏伟的通天塔，那以后还有什么事干不成呢？一定得想办法阻止他们。于是他悄悄地离开天国来到人间，变乱了人类的语言，无法交流的人们做鸟兽散，巴别塔的伟念也就轰然倒塌。

圣经中对语言诞生的描述充满了天谴的色彩，我们也知道事实根本就不是这么回事，但语言的差异的的确确对人类社会产生了深远的影响。不同的语言塑造出不同的行为方式和思维模式，这带来了多姿多彩的文化，却也给不同文化之间划出了纵横交错的鸿沟。作为社会性动物，人类生存最基本的需求自然包括沟通和交流，语言的差异却给这个需求平添了诸多隔阂。难道伟大的巴别塔注定只是存在于幻想之中的空中楼阁吗？

令人沮丧的是，眼下这个问题的答案还是“是”。但机器翻译的出现与发展，至少让我们看到了一丝曙光。

机器翻译源于对自然语言的处理。1949年，洛克菲勒基金会的科学家沃伦·韦弗（Warren Weaver）提出了利用计算机实现不同语言的自动翻译的想法，并且得到了学术界和产业界的广泛支持。韦弗的观点代表了当时学术界的主流意见：以逐字对应的方法实现机器翻译。语言作为信息的载体，其本质可以被视为一套编码与解码系统，只不过这套系统

的作用对象是客观世界与人类社会。既然不同语言描述的对象是一致的，其区别就在于读音和字形的不同。因此可以将字/词看成构成语言的基本元素，每一种语言都可以解构为所有字/词组成的集合，通过引入中介语言的方式，把所有语言的编码统一成为用于机器翻译的中间层，从而实现翻译。同样是“自己”这个概念，在汉字中用“我”来表示，在英语中则用“I”来表示，机器翻译的作用就是在“我”和“I”这两个不同语言中的基本元素之间架起一座桥梁，实现准确的对应。

图8-1 沃伦·韦弗

然而乐观和热情不能左右现实存在的客观阻力，从今日的视角来看，这样的一一对应未免过于简单：同一个词可能存在多种意义，在不同的语言环境下也具有不同的表达效果，逐字对应的翻译在意义单一的专业术语上有较好的表现，但在日常生活的复杂语言中就会化为一场灾难。1964年，美国科学院的下属机构语言自动处理咨询委员会通过调查研究，给出了“在目前给机器翻译以大力支持还没有多少理由”的结论，全面否定了机器翻译的可行性，并建议停止对机器翻译项目的资金支持。经过了概念的热炒之后，机器翻译陷入低潮。

但天无绝人之路，进入二十世纪七十年代后，全球化浪潮的出现催生对机器翻译的客观需求，计算机性能的发展则为机器翻译突破了技术

瓶颈，机器翻译重新回到人们的视野之中。这一时期的机器翻译有了全新的理论基础：语言学巨擘诺姆·乔姆斯基（Noam Chomsky）在其经典著作《句法结构（Syntactic Struc-tures）》中对语言的内涵做了深入的阐述，他认为语言的基本元素并非字词，而是句子，而在一种语言中，无限的句子可以由有限的规则推导出来。语言学的进化也对机器翻译的方法论产生了根本性的影响：韦弗推崇的基于字/词的字典匹配方法被推翻，基于规则的句法分析方法粉墨登场。

这里的“规则”指的就是句法结构与语序特点。显然，基于规则的机器翻译更贴近于人类的思考方式：人类通常会把一个句子视为整体，即使对其进行拆分也并不简单地依赖字词，而是根据逻辑关系进行处理。这使得人类翻译非常灵活，即使是不服从语法规则，甚至存在语病的句子都可以翻译得准确无误。基于规则的机器翻译正因为和人类思路一脉相承，刚一诞生便受到众多推崇，似乎成为一劳永逸的不二法门。可理想虽然丰满，现实却依然骨感。基于规则的机器翻译也很快遇到了新问题：在面对多样句法的句子中，并没有比它的前任优秀多少，任何一款翻译软件都不会把“我勒个去”翻译成“Ohmy God”。

基于规则的窘境迫使研究者们重新思考机器翻译的原则。语言的形成过程是自底向上的过程，语法规则并不是在语言诞生之前预先设计出来的，而是在语言的进化过程中不断形成的。这促使机器翻译从基于规则的方法走向基于实例的方法：既然人类可以从已有语言中提取规则，机器为什么不能呢？眼下，基于深度学习和海量数据的统计机器翻译已是业界主流，这个领域的领头羊则是著名的互联网巨头——谷歌。

某种程度上来说，谷歌翻译（Google Translate）是时下大火的阿尔法狗的师傅。诞生于2001年的谷歌翻译起初几年一直在不温不火的状态中挣扎，直到2004年迎来新掌门弗朗兹·欧赫（Franz Och）。

欧赫是个计算机专家，和语言学一点儿不沾边。还在亚琛工业大学攻读博士学位时，他就开发出一个机器翻译系统，这个系统在美国国家

标准和技术研究所组织的第一届机器翻译系统评测中夺魁。在欧赫看来，统计机器翻译的决定性因素永远是数据规模，句法规则知识对系统的作用相当有限（如果不是反作用的话）。独立于语言的算法使得计算机专家无需通晓语言，只需算法就可以得到理想的翻译结果，而谷歌作为搜索引擎所拥有的海量数据规模显然使欧赫如鱼得水。

2005年，谷歌翻译第一次作为参赛选手参加了由美国国家标准与技术研究所主持的机器翻译测评。事实证明这个新来者是成色十足的黑马：谷歌翻译的得分在从阿拉伯语到英语的翻译中领先了第二名将近5%，可别小瞧这区区5个百分点：它意味着研究上5~10年的差距；更可怕的中文到英语的翻译中——机器翻译最难的一个领域，谷歌翻译的得分领先第二名达到17%！这个差距已经超出了一代人的水平。

谷歌翻译是欧赫的杰作，可参赛的另外两个系统也与欧赫密不可分——亚琛工业大学的系统是欧赫博士期间的作品，南加利福尼亚大学的系统则是欧赫做研究教授时开发的。由于时间上的原因，欧赫在谷歌公司的工作不可能比这两个系统有实质性的改进。那么谷歌翻译为什么会比它的姊妹系统高出一筹呢？其原因就在于谷歌的海量数据。足够的数据使欧赫能够在原始算法的基础上增加参数的数量，进而提高翻译的效果。

谷歌翻译的基本原理是通过对大量平行语料的统计分析构建模型，再通过这个模型翻译。生成译文时，需要先在大量人工翻译的文档中寻找模型并进行合理的猜测，再得出恰当的翻译。针对特定语言可供分析的人工翻译文档越多，译文的质量就越高。通过对模型构建算法和数据处理架构的改进，谷歌翻译的效果不断提升，其第一机器翻译系统的评价可谓实至名归。然而已经处于巅峰的谷歌翻译并未止步：2016年，他们推出了全新的谷歌神经机器翻译（Google Neural Machine Translation），实现了又一次的突破。

传统的统计机器翻译采用的仍然是韦弗基于短语的翻译方式：把句

子分成一个个短语和单词进行独立翻译，再将翻译出来的独立短语进行逻辑整理，重组成句。这样的方法在处理同属印欧语系的语言时问题不大，可一旦用于语法逻辑毫不相干的印欧语系与汉藏语系的语言——比如英语和汉语——之间的互译时，语序的确定就变成了一个老大难问题。但神经机器翻译通过采用整体处理的方式解决了这个难题：它将整个句子视作翻译单元，对句子中的每一部分进行带有逻辑的关联翻译，翻译每个字词时都包含着整句话的逻辑。用一个不甚恰当的类比来描述：如果说传统的统计机器翻译是拆得七零八落的全羊宴，神经机器翻译就是最大程度保持原貌的烤全羊。

相比于传统的机器翻译，神经机器翻译以更少的设计工程量提高了翻译质量。更重要的是，神经机器翻译还实现了“零知识翻译”，即在没有先验数据的情况下，让系统对从未见过的语言进行翻译。这在以往绝对是超乎想象的事情。

机器翻译一直被公认为是自然语言处理，乃至整个人工智能领域中最难的课题之一，这不仅仅是科学家们孜孜追求的技术梦想，也寄托着普罗大众实现自由沟通交流的美好愿望。《星际迷航：企业号》曾绘出这样一幅美妙的图景：星舰通讯指挥官佐藤星利用翻译器发明了linguacode矩阵，通过把翻译器集成到星舰人员佩戴的交流别针上或者植入耳朵里，说不同语言的人们就可以进行自由交流。

这就是人工智能时代的巴别塔。

我不是脸盲晚期
——图像识别

邂逅一枝美丽的花儿，却不知道它的名字，它的特征，它的动人花语，这难道不是一件万分遗憾的事？

微软亚洲研究院携手中科院植物所推出“微软识花”app，其智能花卉识别和知识系统将成为你的“花儿百宝箱”。只需拍摄花儿照片或选取手机图库中的花儿图片，由中科院植物所专家标定的庞大花卉数据库将快速、精确地识别花儿，并通过花语、药用价值等信息，讲述关于花儿的小秘密，让你一秒变身识花达人！

图8-2 “微软识花”软件介绍

这是手机软件“微软识花”的官方介绍。这款软件不仅是微软亚洲研究院和中国科学院植物研究所多年学术合作的结晶，也是人工智能在图像识别领域的高水准表现。即便是专业的植物学家也不可能准确地识别出每一种花朵，这要求广博的专业知识和细致的观察能力，可能还需要

长期的训练。对于计算机也是如此：首先要建立由所有花朵实例组成的训练数据集，再使用深度学习技术对机器识别模型进行训练，以达到精确的识别结果。

花朵识别的难度远高于动物品种的识别：小猫小狗的种类数以百计，而目前已被发生的野生植物已达数十万种，大量闻所未闻的新品种还在不断涌现。因此，微软亚洲研究院的研究员们借助中科院植物所提供的260万张花的鉴定照片作为训练集。但计算机不同于人类：虽然它具备人类无法企及的存储能力和计算能力，但却达不到人类的联想想象水平和抽象概括水平。直接应用这数百万图片的训练集能够得到理想的结果，但一定会是个效率相当低下的过程。人类的经验就是这样一味催化剂，具有增强人工智能学习效率的神奇魔力。

传统的深度学习技术遵循自底向上的规律，通过学习底层的高维数据来获取语法信息背后的高层语义表达。人类学习的过程则是自顶向下的过程：从实践中提取理论，反过来再用理论指导实践。如果利用人类智能的认知规律辅助人工智能的学习过程，将自底向上和自顶向下的学习过程加以结合，就可以大大提升深度学习的效率和精度，这也正是“微软识花”开发者们的设计思路。

根据植物分类学的方法，所有绿色植物的分类、鉴别与命名的依据是“界门纲目科属种”的分类系统，系统中每个层次都有其结构特征，子层次之间也存在关联。识花中的精确识别也采用了这种逻辑上层层递进的方式。在深度学习中，机器首先建立起对“科-属-种”基本层级概念的认知，通过引入关于花卉层级结构的外部先验知识，并通过捕捉科属种之间的联系与区别来指导机器学习。

机器识花参考的也是人类人类观察物体、判断物体种类的过程：我们首先会将注意力放在图片的整体轮廓上，以确定图片中是否有花朵存在，并排除掉了花朵本身以外的其他干扰元素。接下来要确定的就是花朵关键的局部特征，通过形状、颜色、结果等局部特征的组合形成对整

体的认识。花瓣层层叠叠，色泽鲜艳亮丽，颇具富丽堂皇之气的花朵十有八九是国花牡丹，而叶长如盾、素净淡雅、亭亭玉立的通常是出淤泥而不染的荷花。

对于机器而言，识别花朵时首先根据花朵的整体特征确定花的“科”，再通过花瓣的分布与形态等细节来确定它归于哪个“属”，最后通过花瓣的颜色和纹理等更加细微的特征来具体判断它属于哪个“种”。开发者根据这样的过程，独创性地设计了视觉多级注意力模型，并结合深层神经网络技术，用于图像的处理与识别。多级注意力模型首先关注物体本身，即自动关注到图片中花所在的区域，再由粗到精对细节部位进行关注，最后对花朵的部位特征进行学习和识别。

在这个过程中，主要技术难点在于部位特征的关注。为解决这个问题，研究员们提取出了深度神经网络中的一些中间信息，这实际上就是从图像中识别出来的模式。这些模式进行归类后，得到的每个聚类都被机器理解为花的某一个部位的集合，这样就构建出了部位检测模型。将由原始图像生成的每个候选框都经过每个部位检测器，而每个部位检测器则会自动检测出这个候选框内最接近这个部位的区域，这样就实现了部位级别的关注。

深度学习中的另外一个问题就是训练数据量的选择。要提高识别的精确度，提升用于训练的数据量是必经之路。出于效率的考虑，识花使用弱监督学习的方式进行数据的训练，通过对训练图像进行标注提升效率。弱监督学习中的“弱”一方面体现在用于机器学习的数据集只有图像级别的标注，即只标注了这张图片上有什么花，但是并没有标注花朵的整体位置，也并没有标注出关键识别区域的位置。另一方面，训练中的标注数据是收到噪声干扰的不精确数据。因此，弱监督学习的方式既可以兼顾到数据质量的不足，又可以保证用于训练的数据量的庞大，最终保证了识花应用的准确性。

但微软的野心显然不只是辨认花朵这么简单。2015年，微软就上线

了一个颜龄识别的机器人网站how-old.net，这个网站可以根据用户上传的照片从面相上分析人物的年龄。此网站一经推出便火爆全球，判断的正确率也很不赖。这也要归功于微软深厚的人脸识别技术底蕴。

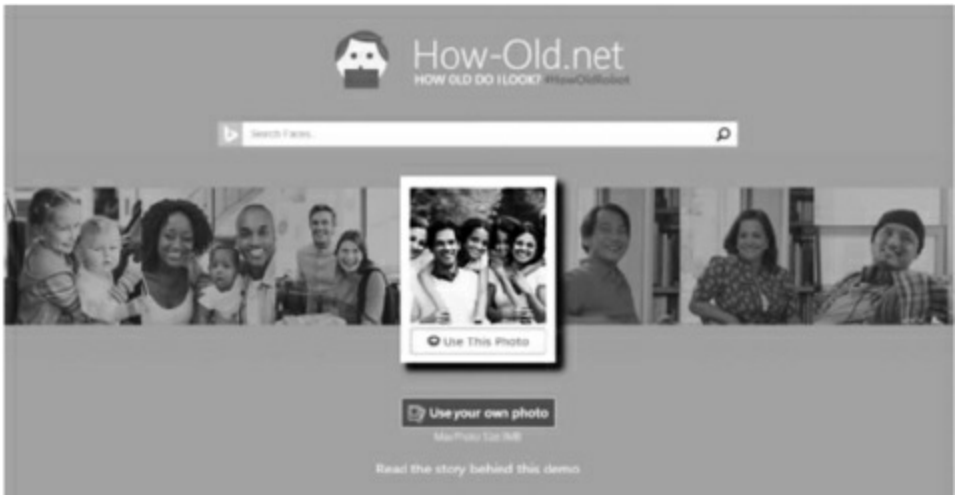

图8-3 微软颜龄识别网站how-old.net

人脸识别是利用分析比较人脸视觉特征信息进行身份鉴别的专门技术，也是人工智能的重要领域，只有在看懂世界的基础上，计算机才能更好地思考。看懂世界的一个重要方面，无疑就是会“认人”。但最近一次人脸识别技术的风头，正是被国内互联网巨头百度抢了去。2017年1月6日，在江苏卫视播出的节目《最强大脑》第四季中，百度公司出品的人工智能机器人“小度”首战告捷：在跨年龄人脸识别任务上，以3:2的比分险胜“最强大脑”代表王峰！

“小度”和王峰的“决战”分为两轮，第一轮，嘉宾章子怡从20张蜜蜂少女队成员童年照中挑出3张高难度照片，选手通过动态录像表演将所选童年照和在场的成年少女相匹配。第二轮，人机共同观察一位30岁以上的观众，随后将他从30张小学集体照中找出。根据节目组的安排，“小度”和王峰第一轮需要识别两个对象。对第一个对象的识别，王峰和“小度”都答对了。第二个对象的识别，现场出现了一个事先没有想到的“状况”：“小度”为一个对象给出了72.98%和72.99%两个极其接近的

匹配答案，这令现场嘉宾大为困惑。可更令人震惊的是，事实证明待识别的对象群组中的确有一组双胞胎！作为科学嘉宾的魏坤琳博士也不禁啧啧称奇。

图8-4 《最强大脑》中的小度

最后，作为“小度”负责人的百度首席科学家吴恩达（An-drew Ng）现场选择72.99%的照片作为最终结果，匹配正确，而王峰则识别错误。正是这一局成为了整场比赛的胜负手：“小度”凭借这1分的优势战胜了人类。

来自百度公司的科学家在现场也对人脸识别技术进行了深入的解读。相比于普通的图像识别，人脸识别的难度更高：每个人的脸部结构特征都是类似的，在判断上就会出现困难。此外，人脸识别中还存在着多种多样的干扰因素：表情、光照、角度等外部因素都会影响到人脸识别的精确度，口罩、墨镜、头发、胡须等人脸遮盖物也会破坏识别效果，更不用说整容、美图或是双胞胎等因素了。

为了达到与人类相似的水平，百度大脑学习包括网上公开的人脸照片、视频影像资料、第三方版权购买内容及一些向大众公开征集的人像照片在内的总共2亿张图片。这个数目相当于看遍了三分之二的美国人

口。但即使这么大的训练样本也有难以解决的问题：在跨年龄阶段人脸识别中，类内变化通常会大于类间变化，这是个巨大的拦路虎。由于跨年龄的训练数据难以收集。基于深度学习的神经网络在数据匮乏的情况下，就难以习得跨年龄的类内和类间变化。

当然办法总比问题多，针对跨年龄人脸识别的第一个难点，百度深度学习试验室的人脸团队选择用度量学习的方法。即通过学习一个非线性投影函数，把图像空间投影到特征空间中。在这个特征空间里，跨年龄的同一个人的两张人脸的距离会比不同人的相似年龄的两张人脸的距离要小。针对第二个难点，考虑到跨年龄人脸的稀缺性。百度深度学习试验室用一个用大规模人脸数据训练好的模型作为底座，然后用跨年龄数据对他做更新。这样不容易出现过拟合的问题。

首次出现在公开场合的百度大脑在人脸识别上的成功，意味着在人工智能领域，我国的科研水平绝不会屈居人后。

穿着白大褂的电脑
——辅助诊断

沃森医生正式诞生于2015年，年仅2岁。他的“哥哥”沃森就是在《危险游戏》中击败人类，勇夺冠军那个万事通。正所谓青出于蓝而胜于蓝，沃森医生是拥有更强专业知识的癌症专家。在印度，它为一名已经无药可救的癌症晚期患者找到了诊断方案；在日本，它只花了10分钟就确诊了一例罕见白血病……

在世界癌症日这天，沃森医生首次在中国坐堂出诊，不过是作为几位人类医生的助手。在天津市第三中心医院的诊室里，收到一位胃癌局部晚期患者的治疗史、分期特征、转移位点、危重病情况等病理数据后，沃森医生“思考”了不到10秒钟，在电脑屏幕上开出了一张详细的西医诊疗方案分析单。在这张分析单里，沃森医生提出的最佳诊疗方案是通过化疗将肿瘤缩小后再进行手术：这和资深肿瘤科医生的判断完全一致。

10秒钟的时间能够做很多事情，对于沃森医生更是如此。在浩瀚的网络空间中，沃森医生根据患者的指标数据，在庞大的数据库里检索了300余份权威医学期刊和200余种医学教科书，以及1500多万网页资料中的关键信息。除了治疗方案，沃森医生还列出了详细的用药、治疗建议以及参考文献全文等内容，甚至在返回中国这位患者身边之前，它还顺便把诊疗方案中的部分内容译为中文。短短2个小时的时间里，20多位癌症患者在沃森医生的协助下获得了诊疗方案，其超高的诊疗效率和精准的诊疗结果让人类医生刮目相看。

经验丰富的人类医生在诊断病症时，依据的是自己多年的临床经

验，沃森医生依据的则是目前全球范围内对相关病例的大数据分析，这种诊断方法被称为循证医学（Evidence-Based Medi-cine）。根据循证医学创始人之一大卫·萨基特（David Sackett）教授的观点，循证医学核心思想是在医疗决策中将临床证据、个人经验与患者的实际状况和意愿三者相结合来制定医疗措施。临床证据主要来自大样本的随机对照临床试验、系统性评价和荟萃分析，这正是大数据处理的用武之地。

由于大哥沃森积累了丰富的自然语言处理经验，自2011年起，IBM公司开始训练沃森医生。通过询问，沃森医生使用自然语言处理技术分析病人的病史和病征，再使用大数据处理技术从专业数据库中搜集信息和数据，并迅速给出诊断提示和治疗意见。在美联社记者的亲身经历中，沃森医生为一名虚拟眼疾患者提供了诊断，诊断准确率达到了73%，一个令人满意的结果。随后，美国保健服务提供商Wellpoint公司与IBM签署了合作协议，这是沃森医生的第一份正式工作：他要协助负责复杂病例的护士完成工作，审查医疗服务提供者的医疗请求，并协助肿瘤临床试验。

随后，沃森医生的足迹开始遍及世界：在克利夫兰诊所、纽约斯隆·凯特林癌症中心和缅因癌症治疗中心，他存储和处理海量数据的能力被用于提升诊疗过程的数据与精准度；在全球最好的肿瘤医院——得克萨斯州立大学安德森癌症中心，他将患者的病历、实验室数据和研究数据进行标准化处理，使用高级分析技术进行深度分析；在梅奥诊所，他分析病人记录和临床试验条件，再在知识库中寻找合适的匹配，并参与了一些概念性试验；在泰国康民国际医院，他使用认知计算算法提高癌症治疗质量并协助开发在线诊疗系统；在印度玛尼普尔连锁医院，他为癌症患者提供基于循证医学的个性化医疗服务，并作为专家系统在网络上上线；在日本富国互助寿险公司，沃森医生的出现甚至让超过三分之一的员工失业……

在我国，癌症是疾病死因中的元凶：我国作为癌症高发国，仅2015

年就有近430万新发癌症病例和281.4万例死亡病例，而中国所有癌症的5年生存率仅为30.9%。造成这种现状的重要原因之一就是肿瘤规范诊疗水平的参差不齐：同一个患者在不同医院、甚至在同一医院的不同医生处，都可能得到不同的诊断结果，这无疑会给患者的诊断和治疗带来困难。但这样的问题也不应该归咎于医生的水平：不同的肿瘤患者需要个性化的治疗方案，这就要求肿瘤医生通过追踪文献及时地跟进和获取大量的循证医学证据，形成治疗方案的设计体系。可繁重的临床任务通常让医生无暇分身顾及文献的阅读，沃森医生的认知计算功能则能够让肿瘤的诊疗更加标准化和规范化，有效地解决了这个问题。

在人工智能和大数据进军医疗领域的征途上，沃森医生并不是独行者：2012年谷歌科学比赛的冠军由一位来自美国威斯康星州的高中生，她设计了一种乳腺癌癌细胞的定位算法，算法的训练则是通过对760万乳腺癌患者的医学影像进行机器学习来完成的。当这一算法应用在实际的诊疗中时，其位置预测的准确率高达96%——这惊人的准确率正是760万张样片的功劳，毕竟一个放射科大夫穷其一生能够阅读研究的病例也不会超过十万个。

人工智能不仅能在医学影像分析这类诊断工作上大显身手，还能够代替人类医生使用柳叶刀。目前世界上最先进的手术机器人就是由美国直观外科公司制造的达芬奇手术系统（Da Vinci Surgical System），这个机器人系统早在2000年就通过了美国食品药品监督管理局的认证，正式取得了上岗的资质。达芬奇手术系统由手术台和远程控制终端两部分组成。手术台是手术的具体实施者：这是个具有三个机械手臂的机器人，每个机械臂的灵活性都远远超过人类的手臂，还配有可以进入人体实时拍摄的微型摄像机，这些配置使达芬奇系统可以实施高难度的手术，其创口也非常小。在另一边的远程控制终端上，微型摄像机的拍摄结果被传回，控制系统再根据这些二维图像还原出人体内的高清三维图像，以便医生随时监控手术过程，并在必要的时候加以人工干预。目前共有3000多台达芬奇手术系统工作在世界各地的医疗机构中，总共完成

了超过三百万例的外科手术。

图8-5 达芬奇手术系统

人工智能与人类医生相比具有不少优势：首先，它们在诊断上的准确率很高，而且随着病例的增加，算法的准确度会进一步提升；其次，它们出现失误的可能性非常低，即使是人类医生容易忽略的微小细节也逃不过它们的法眼；最后，手术机器人的稳定性非常好，它们不会因为受到情绪的影响导致水准出现波动。更重要的是，这些智能程序的成本，通常尚不到同等水平人工的百分之一。

作为目前国际上癌症诊断领域最先进的人工智能，沃森医生已经通过了美国的执业医师资格评定考试。它具有人类大脑无法比拟的学习与决策能力，同时还通过最新医学进展的学习不断深化拓展自身的能力。有了沃森医生的协助，人类医生的诊疗能力将得到巨大提升。

据悉，IBM公司已经被重组为7大部门，其中医疗部门的当家花旦无疑就是沃森医生。

知人知面更知心
——推荐系统

在前文中提及，推荐系统正在蚕食我们思考的权利与愿望。它们根据我们的搜索关键词、浏览历史与交易记录来推测我们的喜好，进而推送我们可能愿意为之打开腰包的商品或服务。推荐系统的设计初衷是帮助在线零售商提高销售额，现在却变成了互联网领域竞相争夺的大蛋糕。互联网的飞速发展让数据成为价值的金矿，推荐系统正是借大数据的东风，凭借创新的算法从金矿里开采出实打实的真金白银。

在现实世界里，我们是一个个有血有肉的真实的人，而在推荐系统中，我们是一个个有1有0的二进制的“人”——一串很长的数字。这串数字中存储着你点击的每个链接、你浏览的每个主题以及你消费的每件商品。多年以来，推荐系统正是通过有效地采集和分析这些数据来决定向你推送哪些产品或是服务。早期的推荐系统只根据对象用户的所有行为做出推荐，但随着计算机处理能力的进化和数据的爆炸式增长，协同过滤（Collaborative Filtering）给推荐系统带来了翻天覆地的变化。协同过滤将物品之间的关联性引入评价体系中，达到了更好的精确度。

协同算法可以分为对用户的协同和对物品的协同。用户协同计算的是用户之间的相似度：根据他们对同一物品的打分差异来计算他们之间的距离。如果两个豆瓣用户张三和李四都给电影《小时代》打了三星，说明他们的品位类似，对应一个较小的距离；如果一个打了五星一个打了一星，他们的距离就拉大了。距离相近的用户被划分在同一个邻集中，同一个邻集中的用户就具有较高的相关性。但这种用户关联策略的劣势在于难以有效地形成邻集：不同的豆瓣用户兴趣不同，可能根本没

有交集；好不容易有个共同评分的电影，却是跟风叫好或是扔番茄的商业大片。这种邻集对于推荐系统来说就没有什么参考价值可言了。

相比之下，对物品的协同就靠谱得多。这种算法依据同一用户对不同物品的打分差异来生成物品之间的距离。豆瓣上给波兰导演克日什托夫·基耶洛夫斯基（Krzysztof Kie lowski）打高分的用户很可能也喜欢希腊导演提奥多罗斯·安哲罗普洛斯（Theod-oros Angelopoulos）的作品，同时对迈克尔·贝（Michael Bay）或是詹姆斯·卡梅隆（James Cameron）的商业巨制嗤之以鼻。显然，两位欧洲导演的小众文艺片就处于同一个邻集之中。一对物品之间的距离通常是根据成百上千万的用户的评分计算得出，具有相对稳定的特点，因而推荐系统可以预先计算距离并更快地生成推荐结果。

即便如此，对物品的协同也存在它的问题：死板。它能够发现喜欢同一样物品的人，却不能找出喜欢同一类物品的人。法国印象派大师克劳德·莫奈（Claude Monet）可谓“睡莲专业户”，其《睡莲》系列画作包含惊人的251幅作品！一群莫奈的拥趸很可能喜欢各不相同的睡莲，可协同推荐算法却识别不出他们的共同爱好。

随着机器学习的发展，这个问题的解决方案逐渐浮出水面——将物品的特征简单化、抽象化，将多数复杂的特征转换为少数一般的特征，这个过程被称为降维（Dimensionality Reduc-tion）。降维的概念来源于数学中的矩阵论，它能将一个复杂矩阵用稀疏的特征值表示，因而在通信理论与信号处理中得到了广泛应用，其思想也被借鉴到计算机科学与其他领域的应用之中。

为了理解降维算法，我们还是举个关于打分的例子：如果有人在豆瓣上给《全球通史》打了5星，给《人类简史》和《枪炮、病菌与钢铁》打了4星，给《看见》打了3星，给《冰与火之歌》打了2星，给《幻城》打了1星，系统就会从中勾勒出这个人模糊的肖像：历史爱好者，喜欢读非虚构类作品，讨厌玄幻小说。读者读过的书可能成百上

千，但他选书的标准完全可以用十个以内的特征来描述，降维算法的作用就是从千百本书中提取出这十个特征。通过与这些特征进行比对，推荐算法能够迅速确定这位读者是否会喜欢一本新书。这种更为一般性的呈现使得推荐算法能准确地发现有着相似但不同喜好的用户。

降维算法的应用开启了人工智能对推荐系统的改造。人工智能对推荐系统最深刻的变革，就是不再把推荐系统看作是独立的推荐结果的组合，而是整体的人机交互行为，通过引入时间维度来达到系统和用户的动态优化。

雅虎首页“今日模块”的推荐工作是较早采用强化学习的推荐系统之一，它最初使用的算法是对于每个页面请求，系统以较大的概率选择当前点击率最高的文章推送给用户，而以较小的概率在其他文章中随机选择一篇来推送。这种算法实现简单，但其推荐精度也有限。对这一算法的改进是对每个文章的点击率的估计及其置信区间同时建模，然后每次选择点击率的估计值与其标准差的和最大的那个文章。这个看起来很复杂的描述实质上就是对用户喜欢的文章和没看过的文章进行平衡，引导用户探索未知类别，进而开发新的关注点。这种算法已被证明对系统和用户的长期交互流程优化有显著效果，雅虎新闻和领英网站的广告系统都是基于这种算法实施推荐。在此基础上，深度学习和神经网络也被应用在了推荐系统的设计中，并取得了良好的效果。

在推荐系统上用功最深的企业之一是美国著名的在线流媒体公司Netflix——热播美剧《纸牌屋（Houseof Cards）》和《黑镜》第三季的制片方。2006年到2009年，Netflix斥资百万美元发起了Netflix Prize竞赛，绝对是推荐系统领域最具标志性的事件，这次比赛不仅为推荐系统领域的研究工作广纳贤才，也有效地打通了产学研之间的层层壁垒，使这项技术从学术圈的自说自话深入到商业界的核心腹地。

图8-6 Netflix用户界面

当然，从比赛中受益最大的还是Netflix自身。Netflix新一代推荐系统的承载形式是“会员首页”，约三分之二的视频观看量都是从首页导入的。Netflix之所以敢于使用推荐系统来驱动首页，就是因为对自身推荐技术的高度自信。而这些技术，恰恰来源于Netflix Prize竞赛。

Netflix推荐系统的核心是个性化视频打分，它是针对每个用户给出个性化推荐结果的基础——结合用户以往的观影数据，就可以通过降维得出观影特征，再根据这些特征完成个性化的推送。除了个性化打分之外，龙虎榜和正在流行也是重要的推荐选择。龙虎榜的作用是在个性化推送中进一步优中选优，正在流行则是借助近期趋势来预测用户的行为——在情人节期间推送几部浪漫的爱情片显然是不错的选择。而基于观看历史的推荐则应用了物物协同过滤：计算两部影片的相似度来决定推荐与否。这里的相似度可以从内容角度进行度量，也可以从行为角度进行度量。

Netflix对推荐系统的评价可谓精当：“推荐系统帮助Netflix赢得关键时刻：当一个会员访问Netflix时，Netflix希望能够帮助他在几秒钟之内就找到他感兴趣的影片，以免他去寻找别的乐子。”据估算，个性化推荐系统每年为Netflix节省的费用可达10亿美金，这正是人工智能所带来的商业福利。

第九章　进化，永不止步
——人工智能新趋势

别怂，就是**GAN**
——生成式对抗性网络

刚刚过去的2016年被视为人工智能崛起的元年，阿尔法狗的横空出世让众多互联网巨头纷纷布局人工智能。随着机器的能力越发强大，人工智能也将彻底打破技术交互的格局。未来的人工智能将以什么样的方式发展？我们不妨做个大胆假设。

在谷歌公司中，从事人工智能相关研究的子机构不只有Deep Mind，还有著名的谷歌大脑计划（Google Brain）。谷歌大脑是专注于深度学习研究的一个项目，先后有杰夫·迪恩（Jeff Dean），格雷格·克拉多（Greg Corrado）和吴恩达等大咖加盟。虽然不像Deep Mind风头正劲，谷歌大脑却也在默默开展关于人工智能的研究。2016年10月，谷歌大脑就公开了一个有趣的实验，展现了神经网络的另外一番潜力。

在这个实验中，研究者使用到三个并不复杂的神经网络执行保密通信的任务，根据密码学的惯例，两个合法通信方被命名为Alice和Bob，窃听者则被命名为Eve.同所有的无监督式学习一样，Alice和Bob并不具有任何密码学的先验知识，只有一串秘密的共享密钥用来保护信息。这

样一来，Alice和Bob就需要自己琢磨出一套加密的算法，这套算法还不能直接告诉对方（告诉对方也就等于告诉Eve），只能通过相互间的反馈信息达成共识。

事实上，这种实验设置已经“颠覆”了密码学的常识。在密码学中，保密体制的安全性要么来源于密钥的保密性，要么来源于作为难解数学问题的加解密算法，但无论在对称密码体制还是在非对称密码体制中，加解密算法的执行方式都是公开的。如果合法通信双方不能就加解密的运算达成共识，实现保密通信就是“巧妇难为无米之炊”。可谷歌的研究人员偏偏不信这个邪，他们要做的就是让神经网络自己开发并学习一套加解密的方法。

实验的具体过程基于对称密码体制执行：给定长度为16比特的明文后，Alice根据自行设定的加密方法输出密文并发送给Bob,Bob利用已知的密钥结合Alice发来的密文，自行探索解密的方法，直到解密出正确的明文。窃听者Eve则可以截取到密文，在既不知道加密算法也不知道密钥的前提下，同样需要自己去猜测明文的内容。当然，Bob和Eve从每一次猜测的结果得到的信息不光是猜测的对与错，还包括错误比特的数目。

如果让普通的计算机执行上面的任务，得到的结果肯定是一团浆糊。但神经网络的表现让人瞠目结舌：在最初的猜测中，Bob和Eve的错误率都维持在较高的水平；可在若干轮次的训练后，借助共享密钥的帮助，Bob几乎可以准确无误地猜到明文，但Eve的猜测精度也在逐渐提升。可以想象，如果继续这样波澜不惊地进行下去，Eve破译明文也只是时间问题。

正是此时，没有任何指导的Alice独立作出了决定：对加密算法进行改进！这个决定的结果立刻直观地体现在Bob和Eve的错误率变化上：两者都产生大幅度的上升。但经过短暂的调整后，Bob的错误率迅速下降到0，意味着他适应了Alice的新版加密算法，实现了对明文的破译；

可怜的Eve就没有这么幸运了，她的错误率稳定在了50%左右，意味着她根本没有摸清算法的门道，只是在做二选一的随机猜测。

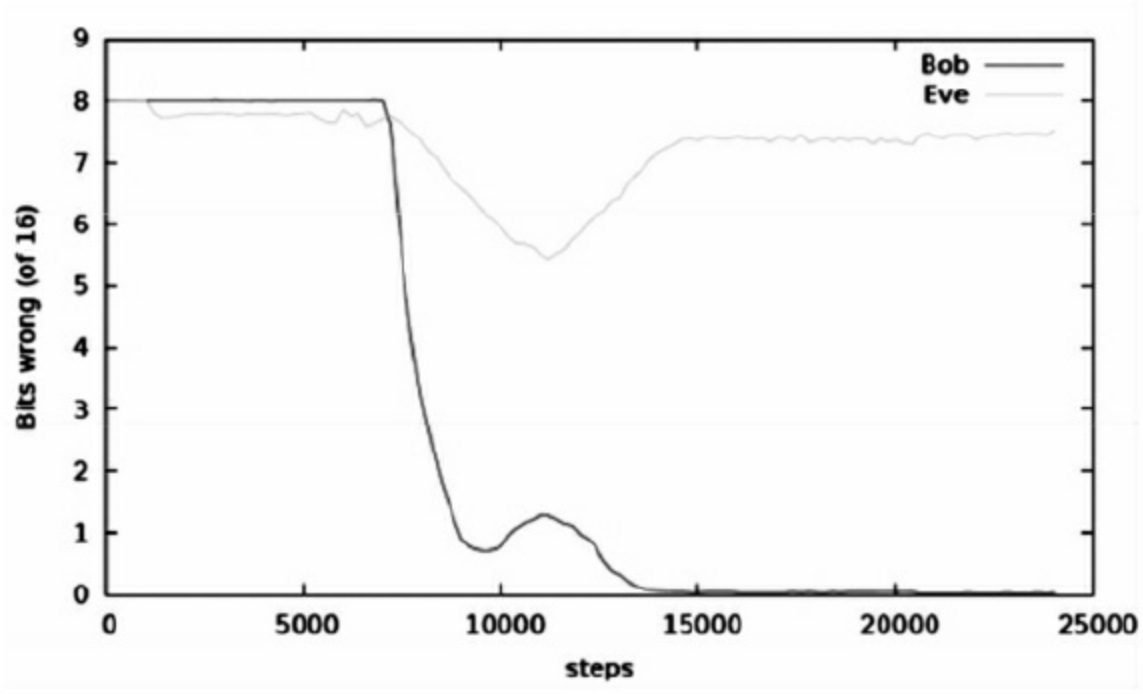

图9-1 谷歌大脑加密实验中Bob和Eve解密结果的错误比特数

来自谷歌大脑的实验再次展示出神经网络的潜能：它们不仅能够在欠缺先验规则的条件下，通过对海量数据的无监督式的学习完成目标，甚至还能够在学习过程中根据实际条件的变化对完成目标的方式进行动态调整。这个实验的环境，在深度学习领域被称为生成式对抗性网络（Generative Adversarial Network），简称GAN.

受博弈论的启发，生成式对抗性网络通过在神经网络之间建立零和游戏框架并引入对抗，来提升学习和进化的速度。2014年，加拿大蒙特利尔大学的伊安·古德菲洛（Ian Goodfellow）与其博士生导师，人工智能大牛约书亚·本基奥（Yoshua Bengio）共同发表了论文《生成式对抗性网络》，正式提出了这一概念。生成式对抗性网络一经提出，便受到了学术界的广泛关注与高度认可，已经成为神经网络与深度学习领域最炙手可热的方向之一。

在生成式对抗性网络中，最核心的概念就是“对抗”，“对抗”的作用则是用来估计生成模型。在对抗的网络框架中，需要同时训练两个模型：生成模型用来获得数据分布；判别模型用来估计训练数据（而非生成模型）中某个样本出现的概率，学习去决定某个样本是否来自模型分布或数据分布。生成模型的训练过程对应为最大化判别模型犯错的概率。生成模型和判别模型之间的对抗可以被类比为造假者与打假者之间的关系：作为造假者的生成模型试图造假，但使用前不检验；作为打假者的判别模型试图检验假货。零和游戏中的竞争促使双方改进方法，直到“假作真时真亦假”，真真假假不可区分为止。

在生成式对抗性网络不同于“对抗实例”。对抗实例是指通过在分类网络的输入上直接基于梯度优化来找出与误分类的数据相似的实例，它不是一个生成模型的训练机制。相反，对抗的实例主要是显示网络行为异常（经常以高置信度将两幅图像分为不同类，尽管两幅图像对人来说是同类）的分析工具。

生成式对抗性网络所要解决的问题是如何从训练样本中学习出新样本，即使再大数量的数据样本也不可能面面俱到，涵盖所有可能出现的情况，所以无论是神经网络还是深度学习，都要解决从已有特征中发现新特征、学习新特征的问题。在这个过程中，之所以选择对抗性学习而非直接学习，原因在于对抗性学习中蕴含着一定程度的监督思想。对抗的引入使学习方式由在已有特征中寻找新特征的蛛丝马迹转化为从已有数据中学习新特征可能出现的方式。这，才是生成式对抗性网络的价值所在。

在实际的应用当中，生成式对抗性网络也得到了大量实用化的改进。加拿大蒙特利尔大学的迈赫迪·米尔扎（Mehdi Mirza）提出了条件生成式对抗性网络的概念，通过增加限制条件让生成的样本更加符合实际预期，有效地抑制了过拟合情况的出现。这里的条件既可以是类别标签，也可以是多模态信息。纽约大学的艾米莉·丹顿（Emily Denton）和

脸书公司（Facebook）的索米斯·钦塔拉（Soumith Chintala）等人合作，将拉普拉斯金字塔（Laplacian Pyramid）结构引入生成式对抗性网络当中，实现了对图像的序列化处理。美国Indico公司的阿莱士·拉德福（Alec Radford）和卢克·梅兹（Luke Metz）则将卷积结构引入生成式对抗性网络之中，将成熟的工程经验应用于新鲜的理论模型，降低了生成式对抗性网络的训练难度。在实际应用中，生成模型通常通过解卷积神经网络实现，而判别模型通常通过卷积网络实现。

生成式对抗性网络如同冬天里的一把火，点亮了深度学习的下一个热点。除了谷歌大脑之外，苹果公司和Deep Mind等新贵也在下大力气开展相关的研究与实验。手机图像软件Prisma已经将生成式对抗性网络应用在图像处理中，用于进行不同风格的创作。它能够生成曲折的线条。识别图像中的阴影，给不同的区域涂上不同的颜色，甚至还能以印象派艺术家的风格完成这些任务。追根溯源，生成式对抗性网络将改变人工智能认知世界的方式，让人工智能以更像人类的方法来学习，是一项具有开创性意义的工作。

人工智能中的负反馈
——强化学习

强化学习（Reinforcement Learning）是近年来深度学习领域迅猛发展起来的一个分支，目的是解决计算机从感知到决策控制的问题，从而实现通用人工智能。具体来说，强化学习是通过智能体在客体环境中采取某种行动，以取得最大化的预期利益。这一概念的灵感来源于心理学中的行为主义理论，即有机体如何在环境给予的奖励或惩罚的刺激下，逐步形成对刺激的预期，产生能获得最大利益的习惯性行为。这个方法在人类的知识获取和技能习得中已经得到广泛应用，在诸如博弈论、控制论、运筹学、信息论、仿真优化方法、多主体系统学习、群体智能、统计学以及遗传算法等其他学科中也有大量的使用。

与传统意义上的深度学习不同，强化学习是基于环境反馈实现决策制定的通用框架，根据不断试错而得到的奖励或者惩罚实现对趋利决策信念的不断增强，强调在与环境的交互过程中实现学习。强化学习与标准的监督式学习之间的区别则在于它并不需要出现正确的输入/输出对，也不需要精确校正次优化的行为。强化学习更加专注于在线规划，需要在探索未知领域和遵从已有知识之间找到平衡。

一个基本的强化学习模型至少包括5个组成部分：（1）环境状态的集合；（2）主体动作的集合；（3）在状态之间转换的规则；（4）规定转换后“即时奖励”的规则；（5）描述主体能够观察到什么的规则。在增强学习中，规则往往是随机的，主体可以观察到的内容是即时奖励和最后一次转换。在许多模型中，主体被假设为可以观察现有的环境状态，这种情况称为完全可观测，否则则称为部分可观测。有时，主体被

允许的动作也会受到限制，就如同我们能支配的财富数目不能超过银行存款的余额。

强化学习的主体与环境基于离散的时间步长相作用。其目标是得到尽可能多的奖励。主体选择的动作是其历史的函数，它也可以选择随机的动作。将这个主体的表现和自始至终以最优方式行动的主体相比较，它们之间的行动差异产生了“悔过”的概念。如果要接近最优的方案来行动，主体就要避免跌入“鼠目寸光”的陷阱，必须根据它的长时间行动序列进行推理。因此，强化学习更擅长解决包含长期反馈的问题。

强化学习的强大能来源于两个方面：使用样本来优化行为，使用函数近似来描述复杂的环境。它们使得强化学习可以使用在更加复杂的环境之中。在很多问题中，模型的环境无法确定，或是根本不存在解析解，所有的知识仅限于给出环境的模拟模型。在这种情况下，从环境中获取信息的唯一办法是和它互动。这些问题都可以通过引入强化学习来解决。

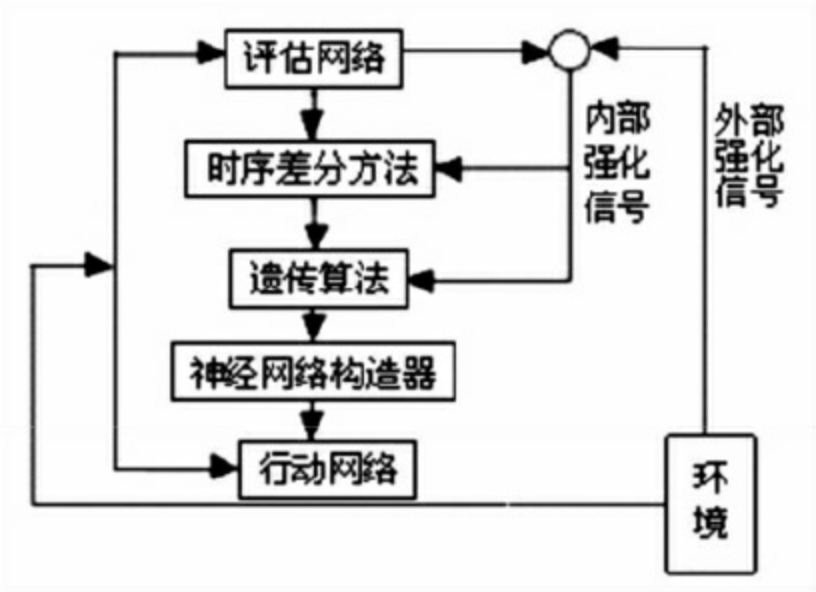

图9-2 强化学习的网络结构

强化学习的迅速发展并不意味着深度学习的退出，两者的关系像是

倚天剑与屠龙刀：强化学习给出了要实现的目标，深度学习则定义了实现目标的方法，两者融合方能得到通向通用人工智能的秘笈。目前最新的深度强化学习算法是Deep Mind公司于2016年11月提出的UNREAL算法（Unsupervised Reinforcement and Auxiliary Learning），这套算法将强化学习和无监督式学习结合起来，并以辅助任务对算法进行改进，是目前效果最好的深度强化学习算法。

和强化学习的概念一样，UNREAL算法的基本思想也借鉴了人类的学习方式。人在完成一项任务的时候，往往会通过并行使用多种辅助任务来实现。一个简单的例子就是需要在微信点赞的时候，我们一方面会给好友群发点赞的消息，也会转发到朋友圈中请别人帮忙扩散，这也正是UNREAL算法的核心。通过设置多个辅助任务同时训练单个行动-评判网络，UNREAL算法可以在加快学习速度的同时进一步提升性能。

UNREAL算法中的辅助任务可以分为三类：控制任务、回馈预测任务和价值迭代任务。具体到图像处理的场景下，控制任务包括像素控制和隐藏层激活控制。像素控制是指控制输入图像的变化使得输出图像的变化最大，因为输出图像的剧烈变化意味着智能体在执行重要的环节，因而通过控制图像的变化能够改善动作的选择。隐藏层激活控制则控制隐藏层神经元的激活过程，尽可能激活量更多的神经元，这就如同大脑开发的程度越高，人就会变得越聪明，也就能够做出更好的选择。回馈预测任务针对的是回馈值无法取得的情况，如果神经网络能够预测回馈值，就会带来更好的表达能力。价值迭代任务的作用是使用历史信息求解损失函数，对价值网络中神经元的参数进行更新，进一步提升算法的训练速度。值得注意的是，虽然UNREAL算法通过不同任务的并行训练来提升神经网络的能力，但其样本数量并未增加，而是在保持原有样本数据不变的情况下，通过对已有数据更加充分的挖掘与利用实现对算法进行提升。

强化学习的一个重要应用领域就是实现目标驱动的视觉导航，简单

来说就是让机器人具备人类的视觉功能。移动机器人的导航问题已有超过半个世纪的历史，其研究内容是实现移动机器人在未知环境中实现从起始位置到目标位置的避障行进。然而在未知环境中，状态空间表现为一个无限维度的连续空间，这让传统的查表类方法不再适用。在这种复杂的条件下，神经网络结合深度强化学习成为解决机器人导航问题的不二法门。

2016年9月，美国斯坦福大学的李飞飞研究组发表了论文《以深度强化学习实现室内场景下的目标驱动视觉导航》，使机器人能够根据实时环境与目标达到指令所指示的任务。论文的作者们使用的是从虚拟迁移到现实的思想：让机器人在高度仿真的环境中执行训练，掌握技能后再迁移到真实场景中，并取得了良好的效果。这将给未来家用机器人的发展带来革命性的变化：连门槛都迈不过去的笨笨的扫地机器人将会成为历史，能够端茶倒水、开门关灯的机器管家即将来临。

机器人导航这个任务在传统的机器人学中是这样实现的：首先要完成对未知环境的建模，这个步骤可以通过机器人上的传感设备采集空间信息来实现；建模完成后就要构建语义地图，也就是在空间模型中为各种物品添加语义信息——这个是椅子，那个是冰箱，诸如此类；接下来就要基于语义地图执行路径规划，机器人只需要根据规划出来的最优路径行进就可以了。当然，我们的描述只用了寥寥百余字，但这百余字背后是百万行数量级的复杂代码，其实现过程非常复杂。

好在强化学习的出现改变了这一切。论文中实现导航的方式就是将目标作为输入，让机器人自行寻找目标，深度强化学习则被应用在机器人具体的训练过程中（具体的技术细节本书受主题限制不做解释，感兴趣的读者可自行阅读论文）。与传统的机器人学方法相比，由于神经网络具有通用性，因此训练出来的机器人掌握的是通用的方法——学会了找椅子，也就学会了找冰箱、台灯、水壶……当然，机器人并没有记住物体的位置，更不知道房屋的结构，但它能够通过不断试错自行寻找通

向每个物体的路径，这也暗合了我国的那句古话：“授人以鱼，不如授人以渔”。

翻不过的那座山
——语义理解

深度学习意味着机器可以越来越多地教会自己如何执行复杂的任务，只有几年前被认为需要人类的独特智慧。自驾车已经是可预见的可能性。在不久的将来，基于深度学习的系统将有助于诊断疾病和推荐治疗。然而尽管取得了令人印象深刻的进展，但一个基本能力依然难以捉摸：语言。像苹果公司的Siri和IBM公司的沃森这样的系统可以遵循简单的口头或打字命令，回答基本问题，但是他们不能对话，也没有真正了解他们使用的话。如果人工智能是真正的变革，这必须改变。

即使阿尔法狗不能说话，它包含可能导致更多语言理解的技术。在谷歌、脸书和亚马逊等公司以及领先的学术人工智能实验室，研究人员正试图通过使用一些相同的人工智能工具（包括深入学习）来解决这个看似棘手的问题，这些工具是阿尔法狗成功的基础，今天的人工智能复兴。是否成功决定了人造智慧革命的规模和性质。这将有助于确定我们是否可以轻松地与机器进行沟通，这些机器将成为我们日常生活中的亲密部分——或者人工智能系统是否保持神秘的黑匣子，即使它们变得更加自主。麻省理工学院认知科学与计算教授乔什-泰恩鲍姆（Josh Tenenbaum）说：“没有办法可以拥有人性化的人工智能系统，没有语言的核心。“这是将人类智慧分开的最明显的事情之一。

应用深度学习语言有一个明显的问题。就是字是任意符号，因此它们与图像根本不同。例如，两个词在含义上可以相似，但含有完全不同的字母；同一个词可以意味着不同的语境中的各种事物。在20世纪80年代，研究人员提出了一个关于如何将语言变成神经网络可以解决的问题

类型的聪明想法。他们表明，词可以表示为数学向量，允许相关词之间的相似性被计算。例如，“船”和“水”在向量空间中是接近的，即使它们看起来非常不同。蒙特利尔大学的研究人员，由本基奥和谷歌公司的另一个组织领导，利用这种洞察力建立网络，使用一个句子中的每个单词来构建一个更为复杂的代表——这是辛顿教授——多伦多大学和一位深入学习研究员——在谷歌公司的工作兼职，称之为“思想向量”。

通过使用两个这样的网络，可以在两种语言之间以极好的精度进行翻译。并且通过将这种类型的网络与旨在识别图像中的对象的网络相结合，可以产生惊人的似是而非的字幕。

坐在位于加利福尼亚州山景城的谷歌繁华总部中心的会议室，公司的研究人员之一，帮助开发这种方法，黎国（Quoc Le），正在考虑一台可以进行适当对话的机器的想法。为什么说话的机器可能是有用的，黎国的野心已经削减了。“我想要一种在机器中模拟想法的方法，”他说。“如果你想模拟想法，那么你应该能够问一台机器它在想什么。”

谷歌已经在教电脑基础知识。今年5月，该公司宣布了一个名为Parsey McParseface的系统，可以查看语法，识别名词，动词和其他文本元素。不难看出，更好的语言理解对于公司来说是有价值的。谷歌的搜索算法用于简单地跟踪网页之间的关键字和链接。现在，使用一个名为Rank Brain的系统，它读取页面上的文本，以便收集意义并提供更好的结果。黎国想要更进一步。他和他的同事使用了在翻译和图像字幕中被证明有用的系统，它们构建了Smart Reply，该回复读取了Gmail邮件的内容，并提出了一些可能的回复。他还创建了一个程序，从谷歌的IT支持聊天记录中学习如何回答简单的技术查询。

最近，黎国建立了一个能够对开放式问题产生通过回应的方案；它是通过从18900部电影进行对话训练的。其中一些答复看起来很厉害。例如，黎国问：“生活的目的是什么？”，方案回应说：“为了服务更大的利益”，“这是一个很好的答案，”他大笑起来。“可能比我的还要

好”。

只有一个问题，当您查看系统的更多答案时，会很快变得明显。当黎国问：“猫有多少腿？”他的系统回答说：“我想，”四点，然后他尝试了一下，“蜈蚣有多少腿？”产生了一个好奇的回应：“八”黎国的节目不知道在说什么。它知道符号的某些组合在一起，但它没有对现实世界的欣赏。它不知道蜈蚣实际上是什么样的，或者它如何移动。它只是智力的幻觉，没有人类认为理所当然的常识。深度学习系统通常可以这样破坏。谷歌创建的用于生成图像字幕的图片将会产生奇怪的错误，例如将街道标志描述为充满食物的冰箱。

黎国问，“生活的目的是什么？”，方案回应说：“为更大的利益服务”。没有人知道如何给机器这些人的技能——如果是可能的话。有什么独特的人类关于这样的品质，使他们超出人工智能的范围？

泰恩鲍姆认为，无论这些网络有多大，今天的神经网络都缺少心灵的重要组成部分。人类有能力从相对较少的数据中快速学习，并且具有内置的三维效能模型。泰恩鲍姆说：“语言建立在其他能力上，其基本能力可能更为基础，这些能力在他们有语言之前就存在于年轻的婴儿中：视觉上看待世界，对我们的运动系统进行行动，了解世界物理学或其他代理人的目标。”如果他是正确的话，那么在机器和人工智能系统中重新创建语言理解并不难于模仿人类学习，心理建模和心理学将是困难的。

诺亚-古德曼（Noah Goodman）在斯坦福心理学系的办公室几乎是裸露的，除了一些抽象画支持一面墙和一些长满的植物。当我到达时，古德曼打电话给笔记本电脑，他的赤脚在桌子上。我们在晒太阳校园漫步，喝冰咖啡。“语言特别之处在于它依赖于很多关于语言的知识，而且还依赖于大量关于世界的常识知识，而这两种知识都以非常微妙的方式进行。”他解释说。

古德曼和他的学生已经开发了一种名为Webppl的编程语言，可以用于给计算机一种概率常识，这在交谈中是非常有用的。一个实验版本可以理解双关语，另一个可以应付夸张。如果被告知有些人不得不等待餐厅的桌子“永远”，那么它会自动确定字面意思是不可能的，而且很可能是挂起了很长时间，并且被惹恼了。该系统远非真正的智能，但它显示了新方法如何帮助人工智能程序以更逼真的方式进行通话。

同时，古德曼的例子也表明，教授语言到机器的困难程度如何。了解“永远”的语境意义是人工智能系统需要学习的东西，但这是一个相当简单和初步的成就。“我想要一种在机器中模拟想法的方法，”他说。“如果你想模拟想法，那么你应该能够问一台机器它在想什么。”

尽管问题的困难和复杂性，研究人员使用深度学习技术来识别图像并在像围棋这样的游戏中表现出色的令人吃惊的成功至少提供了希望，我们也可能处于语言突破的边缘。如果是这样，这些进步会及时到来。如果人工智能作为一种无处不在的工具，人们用来增加自己的智慧和信任，以无缝协作来接管任务，语言将是关键。尤其是人工智能系统越来越多地使用深度学习和其他技术来自己编程。

事实上，随着人工智能系统越来越复杂和复杂，很难设想我们如何在没有语言的情况下与他们进行协作，而无需向他们提出问题，“为什么？”除此之外，与计算机毫不费力地沟通的能力将使他们无限更有用，它会感觉到没有什么奇迹。毕竟，语言是我们最有效的方式来体现世界，并与之互动。这是我们的机器赶上的时间。

（注：本文英文原文为刊载于《麻省理工学院技术评论》2016年9-10月刊的文章Creatingmachinesthatunderst and language is AI'snextbigchallenge.本文为使用谷歌翻译得到的中文译文，有删节。读者可借由此文领略人工智能在自然语言处理领域的最高水准。）

图9-3 本文原文的二维码

定义意念的力量
——脑机接口

2014年6月12日，万众瞩目的第20届国际足联世界杯在巴西城市圣保罗隆重开幕。开幕式上，一名叫做朱利亚诺·平托的下肢瘫痪青年利用“意识”的力量驱动仿生外骨骼，为巴西世界杯进行开球。这次足球历史上最具创意的开球方式，依赖的正是脑机接口技术（Brain Computer Interface）。

图9-4 2014巴西世界杯开球瞬间

所谓脑机接口是在人脑和神经系统与外部设备间创建的直接连接通路。单向脑机接口允许计算机或者接受脑传来的命令或者发送信号到人脑，但两者不能同时进行，双向脑机接口则允许人脑和外部设备间的双向信息交换。

自20世纪90年代中期以来，从实验中获得的此类知识呈显著增长，脑机接口也逐渐从科幻小说中走入现实。在多年来动物实验的实践基础上，应用于人体的早期植入设备被设计及制造出来，用于恢复损伤的听觉、视觉和肢体运动能力。研究的主线是大脑不同寻常的皮层可塑性，它与脑机接口相适应，可以像自然肢体那样控制植入的假肢。在当前所获取的技术与知识的进展之下，脑机接口研究的先驱者们可令人信服地尝试制造出增强人体功能的脑机接口，而不仅仅止于恢复人体的功能。

脑机接口的诞生来源于一次恶作剧般的尝试：1963年，英国伯顿神经病学研究所的威廉·沃尔特（William Walter）医生和他的癫痫病人们开了个小玩笑。当时，脑电技术是辅助治疗的重要手段，具体的实现方式是将电极植入接近大脑皮层的位置，以记录神经细胞活动时所产生的场电势。不走寻常路的沃尔特医生灵机一动，使用电路装置将病人大脑皮层中的场电势信号经过放大后转化为幻灯片的控制信号。这样一来，每当病人产生换片的想法时，无需按动按钮，脑电波产生的电势变化就自动完成了幻灯片的切换，达到了“意念控制”的神奇效果！这是脑机接口技术的第一次完整实现：采集大脑神经信号，翻译转换后控制外部设备。

这次不经意的尝试开启了一扇对人脑研究的新窗口，脑机接口的研究也沿着面向运动功能和面向感觉功能这两大方向逐渐展开。面向运动功能的脑机接口通过发展算法重建运动皮层神经元对身体运动的控制，面向感觉功能的脑机接口则帮助修复受损的听觉、视觉或前庭感觉。

经过一段时间的沉寂后，脑机接口研究在2005年重新回到主流学术界的视野中，关注的焦点则在于脑机接口的实现方式。当时的主流是非侵入式脑机接口（Non-Invasive BCI），这种技术使用的是没有创伤的头皮脑电，其装置方便佩戴于人体，但是由于颅骨对信号的衰减作用和对神经元发出的电磁波的分散和模糊效应，记录到信号的分辨率并不高。这种信号波仍可被检测到，但很难确定发出信号的脑区或者相关的

单个神经元的放电，需要模式识别技术加以辅助。另一方面，另一类侵入式脑机接口（In-vasive BCI）也取得了很大的进展：此类脑机接口通常被通过外科手术直接植入到大脑的灰质当中，以获取高质量的神经信号，用来重建视觉等特殊感觉以及瘫痪病人的运动功能。

除了以上两种技术路径之外，基于细胞培养物的脑机接口也在不断发展。细胞培养物的脑机接口是人体外的培养皿中的神经组织和人造设备之间的通讯机制。这方面研究的焦点是在半导体芯片上培养神经组织，并且从这些神经细胞记录信号或对其进行刺激，建造具有问题解决能力的神经元网络，进而促成生物式计算机。基于细胞培养物的脑机接口已被证明是科学家研究神经元学习、记忆、可塑性、连通性和信息处理背后的基本原理的宝贵工具。

培养的神经元通常通过计算机连接到真实的或模拟的机器人组件，研究人员可以在一个现实的环境中深入研究学习和可塑性，神经元网络能够与其环境相互作用并且至少接收一些人造感觉反馈。这类技术还有助于揭示人类的脑在细胞层面上如何学习特定的计算任务。生物系统在多空间和时间尺度上都有很强的互动性和连接性，以高空间和时间分辨率监控大量神经元活动对于破译神经活动规则、理解神经网络、对大脑的运算过程进行反向工程以提高电脑在模式识别等方面的能力至关重要。

脑机接口技术的发展为机器人的进一步进化提供了可能：实时捕捉大脑复杂神经信号，并用来直接控制外部设备，使得人和机械可以作为一个生命不同组成部分而共存，让机器人真正成为人的意志的延伸。在理想情况下，脑信号可以取代语音等方式，直接用于机器人的操控，实现高效的人机交互与协同。

当然在实现脑机接口操控的过程中，一个难点就是从脑波模式中破译用户的意图。每个人都有独一无二的大脑，这让解密脑波的任务变得更加困难。目前，人与机器的交互方式是人类大脑来适应机器接口，鼠

标，操纵杆，键盘都是静态的设备，用户能够适应并学习使用它们。但要让机器人学会识别脑波，这个过程就要困难得多。

随着脑机接口技术的发展，相关的伦理学争议也渐渐浮出水面，其中最著名的莫过于美国哲学家希拉里·普特南（Hilary Putnam）所提出的“缸中之脑（Brainina Vat）"。当未来的技术条件发展到能够在体外用电信号复制出大脑时，这台“电脑”就有能力模拟你的日常体验。如果这确实可能的话，你要如何来证明你周围的世界是真实的，而不是由一台电脑产生的某种模拟环境？

第十章　这里群星璀璨
——人工智能英雄谱

游刃有余的跨界大牛
——司马贺/西蒙

人工智能注定将和蒸汽机、电力、计算机一样，作为革命性的成果被载入人类发展的史册：要么因推动人类社会完成跨越式变革而流芳百世；要么因挖掘人类的坟墓而遗臭万年。无论如何，那些为人工智能技术奠基的科技大师都值得我们尊敬，作为新时代的图灵和冯诺依曼，他们的探索精神正是驱动人类不断进步的不竭动力。

看到司马贺这个名字，您会想到什么？司马迁，司马光或是司马相如？事实上，这个名字的所有者可是位如假包换的美国人，在血统上和中国没有半点关系，他的英文名字叫做赫伯特·亚历山大·西蒙（Herbert Alexander Simon）。西蒙不仅能熟练使用中文进行读写，在中国访学期间还为自己入乡随俗地起了“司马贺”这个谐音的名字，让人印象深刻。

事实上，这位老先生与中国颇有渊源：他曾先后10次来中国访问交流，中国是除了母国美国以外他生活时间最长的国家。1972年，“乒乓外交”这一历史性的破冰之旅后，司马贺以计算机科学家代表团成员的身份首次来华，为新中国计算机事业的发展献计献策。八年后的1980

年，司马贺又作为心理学家代表团的成员再次来华，正是这次访问让司马贺与中国结下了不解之缘。1983年到1987年间，作为中美科技交流委员会美方主席的司马贺每年都要来中国进行讲学，与中国学术界建立了长期合作关系，在古汉语的计算机翻译、汉字的短时记忆、问题解决和样例研究等问题上都进行了卓有成效的研究，并于1986年出版了学术著作《人类的认知——思维的信息加工理论》。司马贺卓越的工作在中国得到了广泛认可：他是北京大学、天津大学、中国科学院管理学院等单位的名誉教授，1985年被聘为中国科学院心理研究所名誉研究员，1995年当选中国科学院外籍院士。

除了与中国的不解之缘，司马贺的学术成就同样令人叹为观止：他在近四十年的学术生涯中获得空前绝后的9个博士学位！这其中包括1943年他在美国加州大学伯克利分校攻读取得的哲学博士，也包括美国耶鲁大学、瑞典伦德大学、加拿大麦吉尔大学、荷兰鹿特丹伊拉斯谟斯大学和美国哈佛大学等名校因其科学成就而授予的荣誉博士学位。

博士学位的数量还不足以说明司马贺在学术界的地位，研究领域广泛的他在科研上堪称获奖专业户，大大小小的奖项拿到手软。这些奖项中分量最重的莫过于1969年美国心理学会颁发的杰出科学贡献奖、1975年美国计算机协会颁发的图灵奖和1978年挪威诺贝尔委员会颁发的诺贝尔经济学奖——要知道，这可是计算机科学和经济学两个风马牛不相及的领域啊！截至2016年，这位通才式的学者在人工智能和认知心理学领域仍然是被引用次数最多的作者，名下有近一千篇高被引用的论文——搞过科研的人都明白这是多么了不起的成就。

图10-1 美国肖像画家理查德·拉帕波特（Richard Rappaport）所做的司马贺画像

实打实地讲，作为研究者的司马贺把主要精力投入到心理学和管理学的研究当中，人工智能充其量只是他的一门副业，可深厚的数学基础和计算机程序设计功底让司马贺即使在这门副业上同样取得了常人难以企及的成就。作为达特茅斯会议的与会者，司马贺见证了人工智能这门学科的诞生，也在人工智能的发展过程中打下了自己的烙印。

司马贺对人工智能的第一个贡献就是创立了符号主义，这在后来也演变为前文提到的符号主义学派。符号主义认为知识的基本元素是符号，最原始的符号就是物理客体。智能的基础依赖于知识，研究方法则是用计算机软件和心理学方法进行宏观上的人脑功能的模拟。司马贺借鉴了心理学的研究方法，把人脑看成是一个实现信息加工目的的物理符号系统，物理系统表现智能行为必要和充分的条件是它是一个物理符号系统，这也成为传统人工智能的理论基础，这个指令系统中有若干基本操作，实现的功能就是识别不同的模式。在符号主义的基础上，司马贺

等又进一步提出了纯认知系统模型，在原有物理符号系统上增加了情感、认知等人类思维特有的影响因素，完善了符号主义对人工智能的认识。此外，司马贺还提出了组块理论，类比搭积木的方式将零散的构件组成有意义的信息加工单元，来模拟人类思维对信息进行组织或再编码的过程，这一理论在今天的自然语言处理中仍有应用。

司马贺对人工智能的第二个贡献是开创了决策理论。其实“决策”这个概念出自于司马贺的老本行——管理学，也是司马贺对管理学的主要贡献之一。当决策理论被创新性地应用在了人工智能之中时，司马贺提出了决策是管理的核心的论断，以数学方式定义了决策的四个阶段——信息活动、设计活动、选择活动、审验活动，提出了基于“满意度”而非“最优化”的决策模型，并把决策类型划分为程序化决策和非程序化决策两种。在关于决策的一系列贡献中，基于满意度的模型对人工智能产生了重大的影响。在今天的人工智能中应用的海量参数的启发式算法中，在问题的最优解难以求得的情况下，都使用满意度作为评价标准。另一个对人工智能影响深远的理论是有限理性理论，有限理性理论描述了不含情感因素的计算和思考方式，已经成为蚁群、蜂群等智能体行为研究的理论基础。

司马贺对人工智能的第三个贡献是提出了学习模型。司马贺对学习给出了一个简单的定义：性能的改进就是学习。在此基础上，司马贺领导研究了多个知识发现系统，力求从长期积累的物理学、天文学、化学等自然科学学科中的实验数据里面重新发现隐藏的科学定理。这样的学习模型在今天的地球物理学、生物信息学、乃至大数据的处理与价值挖掘中都存在潜在的应用价值。

除了理论贡献之外，司马贺还领导开发了人工智能的一系列实际产品，其中的代表作就是逻辑理论家（Logic Theorist）和通用问题求解器（General Problem Solver）。逻辑理论家由司马贺和他的学生纽厄尔共同发明，通过问题分解和代入的方式来实现机器对数学定理的证明，是

数学机械化进程中的重要成果。通用问题求解器则利用“手段-目的”式的分析方法进行启发式搜索，通过类比对问题进行聚类，进而提出解决方案，这一算法也最终进化为认知领域知名的SOAR系统。

在产品的设计中，司马贺在编程上也小露了一手。在开发通用问题求解器的过程中，他与同事纽厄尔和克里夫·肖（Cliff Shaw）共同发明了信息处理语言（Information Processing Lan-guage）。信息处理语言在超级简明的汇编风格语法中，第一次引入了与人工智能和现代编程语言密切相关的表处理结构，这种结构也为后来高级程序设计语言——诸如C,C++，Java等——奠定了基础。

关于司马贺另外一件不得不说的轶事就是他关于机器下棋的论断。除了科学造诣，司马贺琴棋书画也是样样精通，维基百科上关于他的词条中使用的就是一幅自画像。身为国际象棋高手的司马贺自然对机器下棋情有独钟，在符号主义取得了逻辑理论家和通用问题求解器等进展后，司马贺于1957年作出了一个大胆的论断：计算机下棋的水平可以在10年内超过人类！

遗憾的是，这次司马贺高估了人工智能的水准。计算机下棋的水平确实超过了人类，只不过是在四十年和六十年之后，远远超出司马贺的预期。2001年去世的司马贺见证了深蓝的成功，却无缘目睹阿尔法狗的胜利。但不管怎样，该来的总会到来，今日人工智能的迅猛发展正是对伟大先驱们最好的慰藉。若司马贺泉下有知，想必也会为阿尔法狗而欣喜不已。

得饶人处且饶人
——明斯基的一点过往

有人的地方就有江湖，这是不变的真理。江湖不仅存在于剑客豪侠的刀光剑影之中，也存在于文人墨客的口诛笔伐之下。一句“文人相轻”，道出了这个江湖的凶险。科学界的江湖与侠义道的江湖别无二致，如果说有区别的话，那就是这些高级知识分子的手段更加凶狠，正所谓“软刀子杀人不见血”。

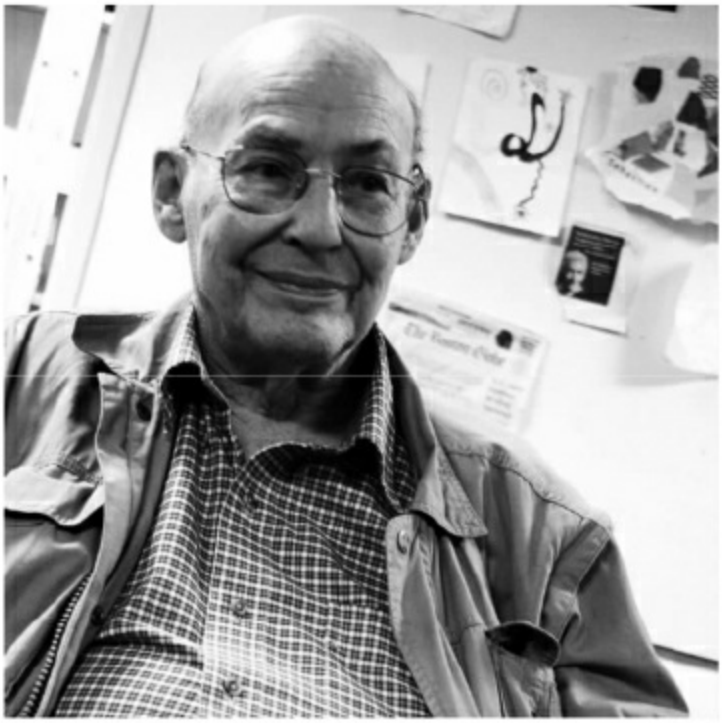

图10-2 马文·明斯基

2016年1月24日，达特茅斯会议的与会者之一马文·明斯基于美国波士顿去世，享年88岁。明斯基作为人工智能的先驱者，创建了世界上第

一个人工智能实验室——麻省理工学院人工智能实验室，并为人工智能领域带来了第一座图灵奖（1969年）。他对具备感知器的机器人触手的研发推进了现代机器人学的进展，他对神经科学与认知科学的研究填补了社会化人工智能的空白，他还为斯坦利·库布里克（Stanley Kubrick）的经典科幻电影《2001：太空漫游（2001:ASpace Odyssey）》客串过科学顾问。当这位百科全书式的科学家离我们而去后，人工智能的第一代研究者也就此谢幕，空留给我们无尽的怀念与追忆。曾与明斯基共事的计算机界巨擘阿兰·凯（Alan Kay）精当地评价了明斯基的一生：

“马文是为数不多的人工智能先驱之一，他用视野和洞见，将计算机从一部超强加法器的传统定位中解放出来，并为其赋予了新的使命——有史以来最强大的人类力量倍增器之一。”

今天，明斯基作为连接主义学派与人工神经网络的代表人物而广为人知，其在1969年发表的经典著作《感知机（Percep-trons）》也是人工智能领域的经典著作。可是在明斯基等身的荣耀背后，隐藏的却是关于神经网络发展中一段并不光彩的往事，一位学术界的青年才俊甚至为此付出了生命的代价。

模拟神经网络的概念诞生于1943年，生理学家麦卡洛克和数理逻辑学家皮茨发表了论文《神经活动中内在思想的逻辑演算》，不仅提出了神经网络的雏形，也成为维纳创建控制论的灵感来源。但提出之后神经网络的研究就陷入了沉寂，直到1957年，一位初出茅庐的年轻人搞了个大新闻。

弗兰克·罗森布拉特于1956年在美国康奈尔大学获得了博士学位，年仅28岁，志得意满的他开始了关于认知系统的研究。仅仅用了不到一年的时间，他的工作就取得了重大突破：他建立了一个名为“感知机（perpectron）”的神经网络模型，并在IBM-704计算机上对这个模型进行了模拟实现。感知机的过人之处在于运行规则遵循生物学的原理，在此基础上可以完成一些简单的视觉处理任务。此后，罗森布拉特又在理

论上证明了单层神经网络在处理线性可分的模式识别问题时的收敛性，用实验证明了感知机具备一定程度的学习能力。1962年，罗森布拉特将他的研究成果整理成一部专著，取名为《神经动力学原理：感知机和大脑机制的理论》出版。这部学术著作的出版一时间引得洛阳纸贵，包括《纽约时报》和《纽约客》在内的诸多媒体都对罗森布拉特的研究给出了高度评价。

罗森布拉特的成果自然也得到了政府机构的关注，美国国防部和美国海军的研究所都资助了他的研究工作，让他获得了充裕的研究经费。突如其来的成功让年仅34岁的罗森布拉特有些忘乎所以，原本内向的他逐渐适应了聚光灯下的生活，他开起了最时髦的跑车，在公众面前大秀钢琴技艺，面对媒体的镜头与话筒侃侃而谈。这幅派头像极了今日频频于公众面前慷慨激昂的某些我国明星学者，只不过罗森布拉特的肚子里确实有货，这些活动虽然对于韬光养晦已久的他来说只是对“人不轻狂枉少年”的一个注脚，却成为他悲剧命运的开始。

图10-3 弗兰克·罗森布拉特的感知机

这时就轮到本节的主人公明斯基出场了。其实明斯基和罗森布拉特

渊源颇深，两人曾共同就读于布朗克斯科学高中（The Bronx High School of Science）。明斯基是1945年毕业生，而罗森布拉特是1946年毕业生，在学制四年的美国高中中，两人至少会有两年的交集。曾经的往事因当事人的离去已不可考，但二人之间的恩怨可能在青年时期就已埋下伏笔。在1960年代，明斯基作为达特茅斯会议的与会者之一，可谓声名显赫。不知不觉间被自己的小师弟抢了风头，这个小师弟又如此张狂，不被放在眼里的大师兄相当不爽。

像明斯基这般的人物自然不会在乎虚名，真正让他不爽的是利益问题，也就是科研经费的分配。工作上，明斯基领衔的麻省理工学院的人工智能实验室也在向美国国防部和海军申请相关经费，但风头正劲的罗森布拉特却把真金白银也抢得一干二净。这让本就反感罗森布拉特做派的明斯基是可忍孰不可忍，于是他想出一招釜底抽薪的妙计。

在关于神经网络能否解决人工智能的问题上和罗森布拉特吵了一架后，明斯基放出了大招：他和麻省理工学院人工智能实验室的同僚，南非学者赛穆尔·帕佩尔特（Seymour Papert）合著了《感知机：计算几何学》一书，证明了单层神经网络不能解决逻辑运算中的异或问题。异或是最基本的逻辑问题之一，如果这个问题都解决不了，那神经网络跟摆设也没有什么区别了。

如果这种讨论仅限于科学问题的范围之内倒也无妨，可明斯基的蒙羞之处就在于他夹带了太多对罗森布拉特个人进行攻击的私货。罗森布拉特也意识到感知机在符号处理方面可能存在限制，但他万万没想到“感知机”的缺陷会被明斯基以如此敌意而致命的方式呈现出来：在1969年《感知机》的第一版中（其封面就是象征感知机无能与致命缺陷的双螺旋连通图），明斯基甚至放言“罗森布拉特的论文大多没有科学价值”。由于罗森布拉特在科研圈的人缘也不怎么样，明斯基的成果就像一支冲锋号，引得各路人等纷纷跳出来指摘罗森布拉特，颇有点“墙倒众人推，破鼓万人捶”的意味，而各个政府资助机构也逐渐停止对神

经网络的拨款。心高气傲的罗森布拉特怎能容忍如此的羞辱？1971年7月11日是罗森布拉特43岁的生日，他在马里兰州的切萨皮克湾划船时发生意外淹死，连尸身都没有找到。至于这究竟是意外事故还是以死明志，就留给各位读者去判断了。

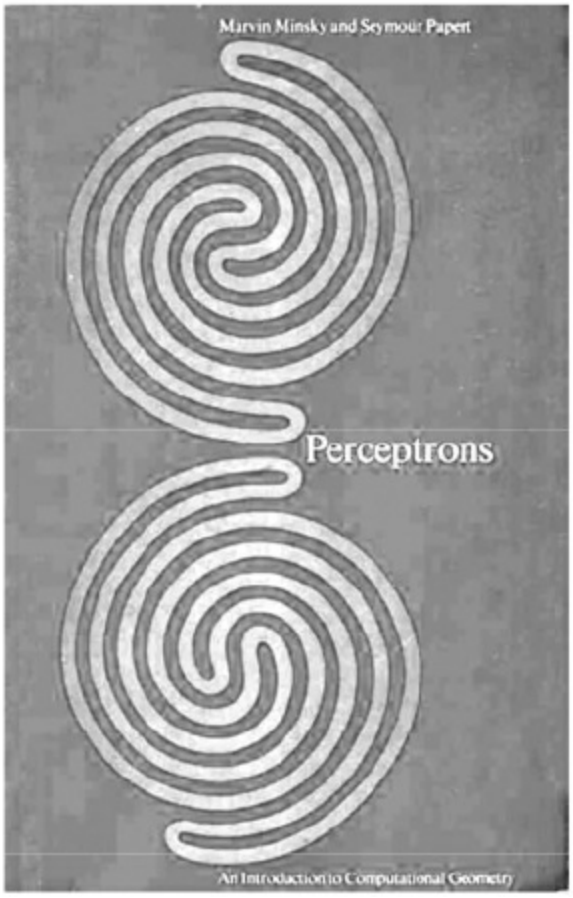

图10-4 《感知机：计算几何学》初版封面

明斯基的本意无非是杀杀小师弟的气焰，却不曾想带来了始料未及的结果。鲁迅先生有诗云：“度尽劫波兄弟在，相逢一笑泯恩仇。”可罗森布拉特再也不能和明斯基度尽劫波，相逢一笑。当20年后《感知机》再版时，明斯基删除了原版中对罗森布拉特进行人身攻击的词句，并亲手写下“纪念罗森布拉特”（In memory of Frank Rosenblatt）。至此，这段不甚光彩的往事也终于可以画上一个句号了。

但这个句号并不意味着神经网络这项技术的终结。明斯基的所指出的感知机的缺陷使神经网络的发展陷入了长时间凛冬般的沉寂，甚至使人工智能这一学科都走入低谷。但在数十年之后，随着新一代天才人物的横空出世，神经网络成功逆袭并取代其他技术一跃成为人工智能的主流。事实证明了罗森布拉特的远见，也还给了含冤而死的他应得的荣耀。2004年，美国电气与电子工程师协会（Institute of Electric and Electronic Engineering）设立了罗森布拉特奖，专门用于表彰在神经网络领域的杰出研究。

基因的力量
——人工智能的救世主辛顿

明斯基对感知机的批判一度让人工智能的连接主义学派走入绝境，在长达20年的时间中无人问津。从今时今日看来，也可以把这段时期理解为黎明前的黑暗，所有的沉寂与蛰伏，都只为等待一位天才人物的出现。

这位天才，就是杰弗雷·辛顿。

辛顿对人工智能的第一个贡献在于将反向传播（Back Propagation）算法引入了多层神经网络的训练当中。虽然小小的异或问题就可以让单层神经网络束手无策，但多层结构的引入就可以使得神经网络能够在理论上执行任何线性或是非线性的函数表达。可多层结构的引入又会带来新的问题：罗森布拉特提出的简单的学习方法无能为力，多层神经网络的训练需要新的算法。

反向传播这一数学方法在20世纪60年代就已经诞生，将其引入神经网络也并非辛顿的首创。但辛顿在前人的基础上更进一步，将反向传播算法发扬光大，才带来了今天人工智能的宏伟蓝图。

20世纪60年代，当全世界的计算机科学家为符号主义的成果而欢欣鼓舞时，彼时的剑桥大学心理学本科生辛顿却认为他们并没有真正地理解大脑。心理学的背景让辛顿敏锐地发觉，神经元的学习方式才是人工智能研究的核心问题。在他的心目中，大脑如同一张全息图，要探索这张全息图的奥秘，研究神经网络是唯一的途径。在多层神经网络中，隐藏层可以发现数据的内在特点，后续层就可以基于这些特点进行操作，

但如何驾驭百万甚至上亿数量级的神经元之间那天文数字一般的组合关系却是无法回避，看起来也无法解决的一道难题。

1978年，辛顿在英国爱丁堡大学获得了人工智能的博士学位，也开始了他宏伟的研究计划：利用计算机硬件和软件来模拟人类大脑，创建一个纯粹的人工智能。这种方式后来被命名为“深度学习”，辛顿也就成为了深度学习的鼻祖。但由于当时计算机软硬件发展水平和训练数据数量的限制，深度学习在诞生初期并没有取得良好的效果，支持向量机（Support Vector Machine）等更加简便易行的算法逐渐占据了学界的主流。但辛顿一直没有放弃自己的信念，他坚信大脑有其独特的运行机制，而大脑的运行机制和当时数字计算机的运行机制完全不同，将给计算机编程的那些套路套用在人工智能上肯定是行不通的。

1985年，辛顿和约翰霍普金斯大学的特里·赛诺斯基（Ter-ry Sejnowski）共同发明了玻尔兹曼机（Boltzmann Machine）。玻尔兹曼机的结构和神经网络类似——由大量基本单元构成的多层网络，但其单元的特性依赖于概率分布而非确定性的决定方式，这一特征显然来源于热力学中的玻尔兹曼分布。这样的特征不仅决定了玻尔兹曼机的数学方法，也决定了它的推理方法：网络中的单元本身拥有能量和状态，学习是由最小化系统能量和热力学直接刺激完成的。当这样的单元聚集在玻尔兹曼机中，它们就形成了一个图形模型：在给定某些代表可见变量的可见单元的已知值后，就可以计算某些隐藏单元的出现概率。这一描述构成了今天绝大多数学习算法的理论框架，玻尔兹曼机的科学意义在于它能够引导出独立于应用领域的一般性学习算法，这种算法会以整个网络发展出的基于周围环境的基础结构的内部模型修改单元之间的联系强度，进而达到更精确的输出（是否想到阿尔法狗呢？）

1989年，辛顿和他的学生燕乐存（Yann LeCun）将反向传播算法应用到神经网络的训练之中，并应用这一算法实现了一个实际任务——将使用反向传播算法训练的神经网络应用于手写邮政编码的识别。五花八

门的手写字体对计算机按规则执行任务的运行模式构成了巨大的挑战，但辛顿的研究应用了来自美国邮政的大量数据资料，证明了神经网络完全足以胜任识别任务，也给深度学习的诞生夯实了基础。

在1992到1993年间，辛顿以极高的效率在一年多的时间里写了200多篇论文，介绍神经网络在学习、记忆、感知和符号处理方法的研究进展。遗憾的是，这些宣传的效果并不理想，直到2004年，关于人工神经网络的研究依然门庭冷落。好在学界的后知后觉并没有影响辛顿的热情，针对业界通过纯粹的监督训练无法构建神经网络的观点，辛顿创造性地提出了非监督训练（Un-supervised Training）的方法，在不知道整个网络应该输出的正确结果的前提下采用逐层训练的方式。随着大数据概念的诞生与处理技术的发展，基于海量数据的无监督学习迅速展现出无与伦比的优势，在其驱动之下的人工智能则一扫往日的阴霾，再次一飞冲天。

以这一方法为代表的深度学习技术迅速取得突破，也让辛顿一鸣惊人。作为目前人工智能领域当仁不让的头号人物，辛顿已经成为各大有志于在人工智能市场上分一杯羹的互联网巨头们与研究机构们争相邀请的座上宾。这正是对辛顿数十年如一日不懈坚守的回报。当然辛顿有的不光是知识，商业头脑也不差：他和两个学生注册了一家从事深度学习业务的公司DNNResearch，整个公司只有他们这三名员工！可就是为了这三个人，2013年谷歌公司愣是豪掷数千万美元，击退了微软和百度等竞争对手，将DNNResearch收入帐下，正所谓“千金易得，一将难求”，这也从侧面验证了培根爵士那句老话：知识就是力量！

当然说到辛顿，就不能不说他背后显赫的家族：辛顿的父亲霍华德·辛顿（Howard Everest Hinton）是位英国的昆虫学家，专门研究甲壳虫，在专业领域的成就虽然不如儿子，却也十分了得：因研究墨西哥水生甲壳虫获得了剑桥大学的博士学位，在伦敦自然史博物馆和布里斯托大学供职，一生还高产地发表了309篇学术论文！

图10-5 韩丁与寒春（辛顿像不像他们）

但对于我们中国人来说，更加熟悉的是霍华德·辛顿的两位表亲，杰弗雷·辛顿的两位表叔/姑：威廉·辛顿（William Hinton）和琼·辛顿（Joan Hinton）。这两位中国人民的老朋友都有中国名字：一位叫韩丁，一位叫寒春。韩丁作为农业机械化的专家，曾受聘为联合国粮农组织中国项目专家和中国农业部高级顾问，其反映新中国土地改革的纪实文学《翻身——中国一个村庄的革命纪实》则向西方世界打开了一扇关于中国革命的窗口。韩丁的妹妹寒春则是第一位获得中国绿卡的外国人，曾经参与曼哈顿计划的寒春是一位优秀的核物理学家，来到中国后在奶牛业和农业机械化事业作出了巨大贡献，2010年于北京辞世。

再往上数三代，辛顿的家世同样声名赫赫：韩丁和寒春的祖母玛丽·艾伦（Mary Ellen）——也就是杰弗雷·辛顿的曾祖母——有个妹妹，叫做艾捷尔·伏尼契（Ethel Voynich），是小说《牛虻》的作者。这部描述意大利革命党人“牛虻”人生经历的作品在欧洲反响平平，却在千里之外

的苏联和中国引起轰动，这位坚强又不失浪漫的革命志士在国内风靡数十年，其英雄形象影响了一代又一代国人。

图10-6 乔治·布尔

艾伦和伏尼契的父亲更是位非凡人物——大名鼎鼎的英国数学家、教育家、哲学家和逻辑学家乔治·布尔（George Boole）。布尔代数、布尔电路、布尔类型、布尔表达式、布尔函数、布尔模型——在各种各种的学科中都能觅得这位大师的踪影。但牛人如布尔也绝对不会想到，他的一位后代会给人工智能这个学科的面貌带来翻天覆地的变化。

所以说基因这个东西，还真是不得不信呢！

从科学家到创业者
——阿尔法狗之父哈撒比斯

当阿尔法狗赚足全世界的眼球时，我等吃瓜群众却忽略了一位重要人物——德米斯·哈撒比斯（Demis Hassabis），正是这位刚刚步入不惑之年的天才发明了阿尔法狗。随着人工智能逐渐占据各大媒体的头条，哈撒比斯也渐渐作为人工智能的新一代代言人，走入了普通人的视野。

1976年7月27日，一位希腊·塞浦路斯血统的父亲和一位中国-新加坡血统的母亲喜得贵子，这个呱呱坠地的婴儿就是哈撒比斯。也许是混血赋予了哈撒比斯基因优势，骨骼轻奇、天赋异禀的他4岁便开始了自己的国际象棋棋手生涯，并多次代表英格兰少年队出战国际比赛。13岁的哈撒比斯就已经达到了国际象棋大师的水准：他的国际等级分是2300分，这在所有14岁以下棋手的等级分中排名第二——仅次于有史以来最优秀的女子棋手，波尔加三姐妹中的小妹朱迪特·波尔加（Judit Polgar）的2335分。

但哈撒比斯的天赋不仅局限于棋盘之上，他用从象棋比赛赢得的奖金买了电脑，开始了自己的程序员生涯：8岁独立编写出计算机游戏，16岁加入牛蛙游戏开发公司（Bullfrog Produc-tions），在游戏《辛迪加（Syndicate）》中负责关卡设计；17岁时又和牛蛙公司的创始人之一彼得·莫里诺（Peter Molyneux）合作，作为主要成员开发出了金手柄奖获奖游戏《主题公园（Theme Park）》。虽然在游戏中取得了与年龄不相符的成功，哈撒比斯却并未被胜利冲昏头脑，清醒的他做出了一个明智的决定——进入剑桥大学学习。

图10-7 德米斯·哈撒比斯

优秀的人在任何领域都会保持优秀，这句话用来形容哈撒比斯再合适不过了。经过四年的学习，20岁的哈撒比斯以计算机科学双重一级荣誉学位从剑桥大学计算机实验室毕业，并重新投入游戏制作行业中，然而这次回头草吃得并不顺利。他建立了自己的游戏工作室，并开发出两款模拟经营类的游戏，却都没有得到大众的青睐，自然也与他的期望相去甚远。痛定思痛的哈撒比斯此时又做出了一个改变命运的决定：退出游戏行业，进入伦敦大学学院攻读认知神经科学的博士学位。

2009年，哈撒比斯在伦敦大学学院取得了博士学位，并继续进行他关于认知神经科学的研究工作。此时他关注的是关于自传体记忆和失忆方面的研究，选择了海马体作为研究对象。海马体主要负责记忆和学习以及空间导向，日常生活中的短期记忆都储存在海马体中，而人类对海马体工作方式的理解至今依然相当有限。哈撒比斯在博士期间最有影响力的工作莫过于创建了关于情景记忆系统的新型理论。在2007年发表的论文《基于建构的情景记忆解构》中，将场景构建视为回忆和想象等思维过程形成的关键步骤。这一理论在认知科学界引发了强烈反响，也入

选了权威期刊《科学》评定的当年十大科学突破。

但哈撒比斯的野望显然不止于此。2010年，哈撒比斯和发小穆斯塔法·苏莱曼（Mustafa Suleyman）以及伦敦大学学院的博士后同事肖恩·莱格（Shane Legg）共同创立了Deep Mind公司。Deep Mind承载了哈撒比斯雄心勃勃的目标：解决智能问题，再用智能解决一切问题。将认知神经科学的前沿理论、机器学习的最新发现与计算机软硬件的技术发展相结合，以实现不限领域的通用用途学习算法，进而实现通用人工智能，哈撒比斯走上了一条从未有人探索的人工智能新路。

Opinion TRENDS in Cognitive Sciences Vol.11 No.7

Deconstructing episodic memory with construction

Demis Hassabis and Eleanor A. Maguire

Wellcome Trust Centre for Neuroimaging, Institute of Neurology, University College London, 12 Queen Square, London, WC1N 3BG, UK

It has recently been observed that the brain network supporting recall of episodic memories shares much in common with other cognitive functions such as episodic future thinking, navigation and theory of mind. It has been speculated that 'self-projection' is the key common process. However, in this Opinion article, we note that other functions (e.g. imagining fictitious experiences) not explicitly connected to either the self or a subjective sense of time, activate a similar brain network. Hence, we argue that the process of 'scene construction' is better able to account for the commonalities in the brain areas engaged by an extended range of disparate functions. In light of this, we re-evaluate our understanding of episodic memory, the processes underpinning it and other related cognitive functions.

Introduction

Episodic memory [1,2], the memory for our everyday personal experiences, is an essential ingredient in shaping how we perceive ourselves [3]. Tulving [2,4] seminally defined three key properties of episodic memory recall: a subjective sense of time (mental time travel), connection to the self, and autonoetic consciousness – a special kind of consciousness that accompanies the act of remembering, enabling one to be aware of the self in subjective time [2]. Others have since identified visual imagery [5], narrative structure [5], retrieval of semantic information [6], and feelings of familiarity [7] as also being important aspects of recollection. Episodic memory recall is, therefore, a highly complex cognitive function that can be conceptually divided into several distinct component processes [3,6,8,9] and is accompanied by a rich recollective experience (Box 1) [2,4]. Although numerous functional magnetic resonance imaging (fMRI) studies investigating the neural basis of episodic memory recall [10–12] have revealed a consistent and distributed network of associated brain regions, surprisingly little is understood about the contributions individual areas make to the overall recollective experience [10–12].

In a stimulating Opinion article published in the February 2007 issue of *Trends in Cognitive Sciences*, Buckner and Carroll [13] make the astute observation that there is extensive overlap in the brain network activated during fMRI studies of remembering the past [10–12], and that engaged during other activities as diverse as thinking about the future [14–16], navigation [17], theory of mind (perspective taking) [18], and the 'default network' [19] (and related to the 'default network', perhaps also 'mind wandering' [20]). This presents the intriguing possibility that these disparate cognitive functions, hitherto treated as distinct, might share common underlying processes. Buckner and Carroll [13] suggest that self-projection might be a crucial common process. They define self-projection as 'the ability to shift perspective from the immediate present to alternative perspectives...requiring a shift in perception from the immediate environment to the alternative, imagined future environment...referenced to oneself'. Thus, their proposal closely connects to Tulving's original ideas of mental time travel of the self [4]. Self-projection is undoubtedly important, perhaps even uniquely so for episodic memory recall and thinking about the future [2]. However, recent convergent neuropsychological [21], electroencephalographic (EEG) [22] and fMRI (D. Hassabis, D. Kumaran and E.A. Maguire, unpublished) findings suggest that at least one further important cognitive function, namely that of richly imagining fictitious experiences [21], is also reliant on the same brain network, but is not explicitly connected to either the self or a subjective sense of time.

Our Opinion article differs from that of Buckner and Carroll [13] in several ways. First we put forward the case for mental scene construction as a well-defined and key component process in supporting recollective experience. Second, we argue that rather than self-projection [13], the process of scene construction is better able to account for the commonalities in the brain networks activated by all the disparate cognitive functions noted above (and possibly others in addition – see Box 2). Third, we link the process of scene construction with existing theories that view the recollection of complex episodic memories as a (re)constructive process [3–5,8,23–25].

Scene construction

We define scene construction as the process of mentally generating and maintaining a complex and coherent scene or event. This is achieved by the retrieval and integration of relevant informational components, stored in their modality-specific cortical areas [26], the product of which has a coherent spatial context [21], and can then later be manipulated and visualized. The full recollective experience of richly recalling an episodic memory [2], for example, remembering what you did last Saturday evening, is nearly always accompanied by complex mental imagery [5] of that event played out within a spatial context [27,28] – likewise if you

Corresponding authors: Hassabis, D. (d.hassabis@fil.ion.ucl.ac.uk); Maguire, E.A. (e.maguire@fil.ion.ucl.ac.uk).

Available online 4 June 2007.

www.sciencedirect.com 1364-6613/$ – see front matter © 2007 Elsevier Ltd. All rights reserved. doi:10.1016/j.tics.2007.06.003

图10-8 哈撒比斯的论文《基于建构的情景记忆解构》

2014年，Deep Mind被谷歌公司以四亿美元的天价收购。后来的事

实证明谷歌并没有白花这笔钱：哈撒比斯和他的团队成员基于深度学习技术先后开发出了能玩游戏机和能下围棋的人工智能，将辛顿二十年如一日为之付出心血的算法的强大力量直观而生动地展现在世人眼前。同时，Deep Mind将深度神经网络与所有动物都有的通过大脑多巴胺驱动奖励机制的“强化学习（Re-inforcement Learning）”结合起来，在神经图灵机、人工智能辅助医疗以及人工智能安全等方向上取得了颇多建树。

在哈撒比斯眼里，真正的人工智能应该是具备以下特征的算法：能像生物系统一般自适应地学习，仅使用原始数据就能从零开始掌握任何任务。而在未来，通用人工智能将与人类专家合作解决一切问题——从癌症的治疗到气候的长期演变、从人类基因组的破解到宏观经济理论在社会中的应用，这些问题的复杂程度已经超出了人类能够处理的极限，却正是机器的拿手好戏。通用人工智能可以帮助人类从庞大的数据量中提取出可用的知识，再从知识中筛选出正确的见解。

关于通用人工智能美好愿景也有其阴暗面，在超级智能诞生之后，人类将何去何从？Deep Mind的所作所为是否代表着潘多拉魔盒的开启？对此，哈撒比斯有自己的看法：“由于这些系统变得越来越复杂，我们需要思考如何充分利用它们，以及它们又能将什么东西进行优化，如何进行优化。”他在接受英国《卫报（The Guardian）》记者伯顿·希尔的采访时说：“技术本身是中立的，但它是一个学习系统，所以不可避免的，它们会承担一些价值体系的印记和设计者的文化，所以我们需要非常小心翼翼地思考这些价值观。”

关于超级智能，哈撒比斯做出了如下的解释：“我们需要确保目标精确详细，并且没有什么模糊的地方，不会随着时间的流逝而发生变化。但在所有的系统中，最顶层目标仍然由它的设计者确定。这可能需要系统自己想办法达成目标，但它不能自行创造目标。看吧，这些是有趣又有难度的挑战。因为这些全新的强大技术需要符合伦理和有责任感地使用，而这就是我们积极呼吁讨论和研究这一事宜的原因，所以当那

个时间窗口到来时，我们能够已经做好了准备。”

当然，在阿尔法狗给通用人工智能带来第一丝曙光的今天，谈论人类命运未免显得杞人忧天。更现实的路线图是将它融入谷歌地图的导航，融入YouTube的视频推荐、融入谷歌翻译的整句处理——总而言之，融入日常的生活当中。人工智能的进步必然会给人类也带来改变，至于这种改变以什么样的方式出现，又会产生什么样的影响，就只有时间能够回答了。可至少哈撒比斯相信，即使是通用人工智能，也终将处于人类的掌控之下。

既然哈撒比斯有理由如此乐观，我们又为什么不呢？

作者简介

王天一

北京邮电大学通信与信息系统专业工学博士，贵州大学大数据与信息工程学院副教授，研究方向为量子通信与物联网/大数据技术。